

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:35:59 ; Search time 43 Seconds

(without alignments)
803.087 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 1887

Sequence: 1 MPLELELCGRWVGQHPHF.....EEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1804	95.6	359	2 JC7321	N-acetylneuraminic
2	619	32.8	337	2 H6432	Spore coat polysac
3	532.5	28.2	341	2 I44651	capsular polysacch
4	510.5	27.1	346	2 I69836	neuB protein - Esc
5	487	25.8	334	2 D81276	N-acetylneuraminic
6	419	22.2	350	2 F97169	sialic acid syntha
7	382.5	20.3	349	2 S60760	polysialic acid ca
8	370	19.6	373	2 S39722	spore coat polysac
9	358	19.0	356	2 D87604	neuB protein, prob
10	356.5	18.9	343	2 B81275	N-acetylneuraminic
11	354	18.8	343	2 H81318	N-acetylneuraminic
12	303	16.1	340	2 B64542	spore coat polysac
13	287	15.2	340	2 H71965	sialic acid syntha
14	226	12.0	378	2 H71307	polysialic acid ca
15	122.5	6.5	331	2 B90173	spore coat polysac
16	119	6.3	352	2 A11899	neuB protein, prob
17	110	5.8	134	2 S53114	N-acetylneuraminic
18	110	5.8	265	2 B75161	N-acetylneuraminic
19	108.5	5.7	351	2 S74729	spore coat polysac
20	108.5	5.7	633	2 D64222	sialic acid syntha
21	108	5.7	63	2 S04974	hypothetical prote
22	108	5.7	338	2 E72388	antifreeze protein
23	107.5	5.7	337	2 A97010	carboxysome format
24	107	5.7	276	2 E72643	phospho-2-dehydro-
25	106.5	5.6	478	2 B70342	carboxysome format
26	102.5	5.4	1140	2 S73786	DNA topoisomerase
27	102	5.4	63	2 S04973	antifreeze protein
28	101	5.4	1107	1 S52517	glutamy1-trna (Gln
29	101	5.4	1175	2 T20346	hypothetical prote
					myosin I heavy cha
					pyruvate carboxyla

ALIGNMENTS

RESULT 1

JC7321

N-acetylneuraminic acid 9-phosphate synthase (EC 4.1.3.-) - mouse

N:Alternate names: N-acetylneuraminic acid synthase homolog

C:Species: Mus musculus (house mouse)

C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 06-Oct-2000

C:Accession: JC7321

R:Nakata, D.; Close, B.E.; Colley, K.J.; Matsuda, T.; Kitajima, K.

Biochem. Biophys. Res. Commun. 273, 642-648, 2000

A:Title: Molecular cloning and expression of the mouse N-acetylneuraminic acid 9-phosphat

A:Reference number: JC7321

A:Accession: JC7321

A:Molecule type: mRNA

A:Residues: 1-359 <NAK>

A:Cross-references: DDBJ:AB041263

C:Comment: This enzyme is a cytosolic enzyme and as a housekeeping enzyme as well, which cetyl-D-mannosamine 6-phosphate to synthesize N-acetylneuraminic acid 9-phosphate.

C:Genetics:

A:Gene: sialic acid

C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpsA

C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 95.6%; Score 1804; DB 2; Length 359;

Best Local Similarity 94.2%; Pred. No. 1e-133;

Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 MPLELELCGRWVGQHPHFIIAIGQHQGDLDAKRMIRMAKCGADCAKFKQSELEF 60

Db 1 MPLELELCGRWVGQHPHFIIAIGQHQGDLDAKRMIRMAKCGADCAKFKQSELEF 60

QY 61 KFNKALERTYTSKHSWGKTYGHEKHLEFSDQYRELQRYAEVGIFFFTASGDEWAVE 120

Db 61 KFNKALERTYTSKHSWGKTYGHEKHLEFSDQYRELQRYAEVGIFFFTASGDEWAVE 120

QY 121 FLHNLNVPFKVGGDTNNPYLEKTAKGRPMVVISGMSQMDTMKVQYQIVKPLNPNFC 180

Db 121 FLHNLNVPFKVGGDTNNPYLEKTAKGRPMVVISGMSQMDTMKVQYQIVKPLNPNFC 180

QY 181 FLOCTSAYPELQEDVNLRVISEYQKLPDPIPIGSGHETGIAISVAALGAKVLERHIT 240

Db 181 FLOCTSAYPELQEDVNLRVISEYQKLPDPIPIGSGHETGIAISVAALGAKVLERHIT 240

QY 241 LDKTWKSGDHSASLEPELAEVRSVRLVERALGSPTKQLLPCEMACNEKLGSVVAKVK 300

Db 241 LDKTWKSGDHSASLEPELAEVRSVRLVERALGSPTKQLLPCEMACNEKLGSVVAKVK 300

QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS 359

Db 301 IPAGTTLTDLTVKVGEPKAYPPEDIFNLGKVKLVLTVEEDDTIMEELVDNHGKKIKS 359

RESULT 2

H64432
 spore coat polysaccharide biosynthesis protein E homolog - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: H64432
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Aron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: H64432
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-337 <BUL>
 A:Cross-references: GB:U67549; GB:L77117; NID:g2826363; PIDN:AAB99068.1; PID:g1591717; T1591717
 C:Genetics:
 A:Map position: REV1006919-1005906
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 Query Match 32.8%; Score 619; DB 2; Length 337;
 Best Local Similarity 40.0%; Pred. No. 7.3e-41;
 Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;
 QY 11 RWVGQHPCTIIAETGQHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEP 70
 DB 9 RYVGKEPTIIAEGGLNHGIDIGKELVKEAKCGADAIKFSQVHT-----D 58
 QY 71 YTSKSGWTKYGEHKKHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFF 130
 DB 59 FISKKS---EYELFKSLSEEEFYELKEAEGIKIMISTPLDKYVDILNKNVPF 115
 QY 131 KVGSGDTNFPYLEKTAKGRPMVSSGMQMDTKQVYQIVKPLNPF-----FLOCTS 186
 DB 116 KTASGLTYPLEKVAKTGPKVILSTGMSDIG---EWEAVKLENNGCRDILLHCIS 172
 QY 187 AYLQPEDNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHTLDKTKW 246
 DB 173 SYPTPYEDVNLNAIKTKSIF-NIPVGSHTLIGILAPVVSVAGADVIEKHFTLDKNME 231
 QY 247 GSDHSLSEFGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGII 306
 DB 232 GPDHLSADPEEFKEWNNIRLVKMLGSGEKTIPMSERDVIVEARSRIVAKNIKKGEY 291
 QY 307 LTMDMLTVKVGEP-KAYPPEDIFNLVGKKVLTVEEDDTI 345
 DB 292 LSVDNISFK--RPGRIETKYLSTIILNRKIKNDKEEDII 329
 RESULT 3
 T44651
 capsular polysaccharide biosynthesis protein cpsN [imported] - Streptococcus agalactiae
 C:Species: Streptococcus agalactiae
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jun-2000
 R:Chaffin, D.O.; Yim, H.H.; Beres, S.B.; Sweet, E.S.; Nittayajarn, A.; Rubens, C.E.
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z22821
 A:Accession: T44651
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-341 <CHA>
 A:Cross-references: EMBL:AF136833; PIDN:AAS3074.1
 A:Experimental source: strain COH1; serotype III
 C:Genetics:
 A:Gene: cpsN
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 Query Match 28.2%; Score 532.5; DB 2; Length 341;
 Best Local Similarity 36.8%; Pred. No. 4.5e-34;
 Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAEIQHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEPPTSXHSWK 79
 DB 3 YIIAIEGCHNGDINLAKWVDVAVSCGVDAVKFQTFKAELKLSIFAPKAEY-QKATTGT 61
 QY 80 TYG--EHRKHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFFKVGSDT 137
 DB 62 ADSQLEMTKRLSESPYELMRDVAISKGVETFTFPDEESLEFLISTDMPTIYKIPSGBI 121
 QY 138 NNFPYLEKTAKGRPMVSSGMQMDTKQVYQIVKPLNPF-----FCFLQCTSAYPELQPE 193
 DB 122 TNLPLEKIGKQKKVILSTGMA---VMEIHAQVNLIRQNTTIDISILHCTEYTPFP 178
 QY 194 DVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHTLDKTKWGSDSAS 253
 DB 179 SLNLNVHTLKDEFKDLTIGYSDHSIGSEVPITAAAAMGAEVIEKHFTLDTNMEGPDHKAS 238
 QY 254 LEPGELAEVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMDMLT 313
 DB 239 ATPDILAAALVKGVRIVEQALGRFEKIPDFVEEKNKIVARKSVVALKPIKKGDYISNIENIT 298
 QY 314 VKVGEP-KAYPPEDIFNLVGKKVLTVEEDDTIMEELVDN 352
 DB 299 VK--RPGNGISPMNWDYILGQAQDDFEDEVIRDSRFEN 336
 RESULT 4
 I69836
 neuB protein - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 20-Jun-2000
 C:Accession: I69836
 R:Annunziato, P.W.; Wright, L.F.; Vann, W.F.; Silver, R.P.
 J. Bacteriol. 177, 312-319, 1995
 A:Title: Nucleotide sequence and genetic analysis of the neuB and neuB genes in region 2
 A:Reference number: I55145; MUID:95113767; PMID:7814319
 A:Accession: I69836
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-346 <RES>
 A:Cross-references: EMBL:U05248; NID:g454079; PIDN:AAC43302.1; PID:g454081
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
 Query Match 27.1%; Score 510.5; DB 2; Length 346;
 Best Local Similarity 36.3%; Pred. No. 2.4e-32;
 Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;
 QY 20 FIIAEIQHQGLDVAKRMIRMAKCGADCAKFOKSELEFKNRKALEPPTSXHSWG- 78
 DB 5 YIVAEIGCHNGSDVIAEMILKAEAGVNAVQTFKADKLISAIAPKAEYQIKNTGEL 64
 QY 79 KTYGEHKKHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFFKVGSDTN 138
 DB 65 ESQLEMTKLEKMDYDLHLMYAVSLNDVFTFPDEDSIDFLASLKQIKWIPSGELL 124
 QY 139 NNFPYLEKTAKGRPMVSSGMQMDTKQVYQIV---KPLNPFCLQCTSAYPELQ 192
 DB 125 NLPYLEKIAKLPIDPKKIISTGMATIDEIKQSVFIPNNKVPVGNITILHCTEYTPFP 184
 QY 193 EDNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHTLDKTKWGSDSAS 252
 DB 185 EDVNLAINDLKKHFPKNNIGFSDHSGGFVAAIAAPVGYITFIEKHFTLDKSGPDHLA 244
 QY 253 SLEPGELAEVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMDM 312
 DB 245 SIEPDELKHLICIGVRCVEKSLGSKSVVTASERKNKIVARKSVIAKTIKKEGVFSEKNI 304
 QY 313 TVKVGEP-KAYPPEDIFNLVGKKVLTVEEDDTIMEELV 350
 DB 305 TTK--RPGNGISPMWYNLLGK-----IAEQDFIPDELI 336
 RESULT 5
 D81276

N-acetylneuraminic acid synthetase (EC 4.1.1.3.-) Cj1327 [imported] - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: D81276
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Cronin, A.; Fraser, H.; Morgan, E.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, S.; Barrett, P.; Whitehead, S.; Woodward, J.; Zhang, X.
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence and a novel virulence factor.
A:Reference number: AB1250; PMID:20150912; PMID:10688204
A:Accession: D81276
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-334 <PAR>
A:Cross-references: GB:AL111168; NID:g96968723; PIDN:CAB73754.1; PID:g9696876
A:Experimental source: serotype OZ, strain NCTC 11168
C:Genetics:
A:Gene: neuB2; Cj1327
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpsA
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match	25.8%;	Score 487;	DB 2;	Length 334;
Best Local Similarity	35.1%;	Pred. NO. 1.6e-30;		
Matches 118;	Conservative 64;	Mismatches 138;	Indels 16;	Gaps 5;

QY 21 IIAEIGQNHQGDLDVAKRMIRMAKECGADCAKFQKSELEFFKNRKALERPYTSHSWG-K 79

Db 6 IIAEAGVNHNGDLNLAKLIEIAADSGADPVKFQSFKAKNICISTKAKAPYOLKTTANDE 65

QY 80 TYGEHKRHLEFSDOYRELORYAEVVGIEFTASGMDMAVERIHEINVPFFKUGSGDNN 139

66 SOLOWOKLEI DI KAKHEL II HAKKNIA EI STODNI PQVDI I NEI CI VIKETI DSCETIN 125

[illegible]

QY 140 F P I E K I A R G R E N W I S S G M Q S M D I M K Q V Y I V --- K P L N P N F C F L Q C T S A Y P L Q P E D V N 198

D6 126 LPYLKXIANKKIIILSTGMANLGEIEEALNVLCNKGAKRQNTLLHCTTEYPAPFNEVN 185

Qy 197 LRWISEYQKLFDPIDPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSASLEP 256

Db 186 LKAMQSLKDAF-KLDVGYS DHTRGIIHSLAAVALGACVIEKFTLDKNMSGPDHKASLEP 244

Qy 257 GELAE LVR SRL VERALGSPTKQLLPCEMACNEKLGSVVAKVKIPECTILTMDMLTVKV 316

Db 245 OELKMLCTOIROIOKAMGDGIKYAKSKECKNINIVRKSIVAKKDKIKKGEIIESEGNITTK- 303

Qv 317 GEPKAYDPBDEIE-----NINGKXVUUVWEEDDNTIME 347

[illegible]

DB -----RPANGISAWRYEEFLGKIATKNYKEDELIRE 334

RESULT 6

F97169 sialic acid synthase [imported] - Clostridium acetabutylicum

C:Species: Clostridium acetobutylicum

C:\date: 14-sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:\Accession: F97169

R.Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, D.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Reference number: A96900; MUID:21359325; PMID:21359325

A/Accession: F9/I69
A/Status: preliminary

A_Molecule type: DNA
A_Residues: 1-350 <KUT>

A; Cross-references: GB: AE001437; PIDN: AAK80145.1; PID: g15025183; GSPDB: GN00168
A; Experimental source: *Electridium acetabulum* AM9004
A; Experimental source: *Electridium acetabulum* AM9004

Experimental source: *Crostidium acetabularium* AlCC824
C:Genetics:
J. Gen. 81201-12

A:Gene: CAC2187
C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis p

Query Match 22.2%: Score 419; DB 2: Length 350.

Best Local Similarity 32.3%; Pred. No. 3.6e-25;
Matches 112; Concentration 73; Mismatch 122

Matches 112; conservative 73; mismatches 132; indels 30; gaps 10

Qy	10	GRWVGQHPCFIIAIBIGONHQOGLDDVAKRMIRMAKECGADCAKQF-----KSELEF	60
Db	9	GKKIGEGQRTIIAEMSAHNQDFRAVEI:IIKAAKRSGADA:IKLQTYTPDITTFDSNEY	68
Qy	61	KFNKALERPPTYKHSWKTYGEHRHLEFSDQYR-ELQRYAEVGVIPFTASGDMEMAV	119
Db	69	FQIKQGITWDGTTTLH--KLXEE-----AYTPWQPKLKEIAEBGLICFSSPFDNTSV	120
Qy	120	BEFLHNLNVPFKVSGDNNPYLEKTAAGKRPWV:ISSGQSMQSDMTWKQYQIV----KPL	175
Db	121	DFLEKMEVPATKVASFEUTDIPFIEYIASKQKPVIMSTGIAEMG---EYEVAACRQOG	177
Qy	176	NPNCFQLQCTSAYPQLQPDVNLRVISEYKQFLPDPIGVSGHETGIATISVAVALCAKYL	235
Db	178	NENVILLKCSSYSPLEEDINLKTIPNWRFAF-NCSVGLSDHTMGYSVATAAVALCATVI	236
Qy	236	ERHITLDTKWKGDHSASLEPEGAELVRSVRLVERALGSPTKQLLPCEMAKNEKLGKSV	295
Db	237	EKHFTLKRSOGPDSAFSMEPEEFSAMVKS:REVEKALGVITYELTE-KQKNSRQHSRSL	295
Qy	296	VAKYKIPBGTTLTMDMLTVKVGEPK-AYPPEDI:FNLVGKKVLVTVEE	341
Db	296	FWXDKIKKGFTFXE--NVKSIRPAFGLTKYIEBVI:GKKARVDVKK	340

RESULT 7

S60760
polyglutamic acid capsule biosynthesis protein SiaC NME0068 [imported] - *Neisseria meningitidis*
C:Species: *Neisseria meningitidis*
C:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 19-Jan-2001
C:Accession: S60760; B81241
R:Edwards, U.; Mueller, A.; Hammerschmidt, S.; Gerardy-Schahn, R.; Frosch, M.
Mol. Microbiol. 14, 141-149, 1994
A:Title: Molecular analysis of the biosynthesis pathway of the alpha-2,8 polysialic acid
A:Reference number: S60758; MUID:95131927; PMID:7830552

A:Accession: S60760
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-349 <EDW>
A:Cross-references: EMBL:M95053; NID:G520732; PID:AAA20477.1; PID:G520735
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1994
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.B.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. A:Reference number: A81000; MUID:2015755; PMID:10710307

A;Accession: B81241

A;Molecule type: DNA
A;Residues: 1-349 <TT>
A;Cross-references: GB:AE002366; GB:AE002098; NID:G7225284; PIDN:AAF40535.1; PID:G7225284
A;Experimental source: serogroup B, strain MC58
C;Genetics:
C;Gene: NMB0068
C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cpsA

Query Match	20.3%	Score 382.5;	DB 2;	Length 349;
Best Local Similarity	30.5%;	Pred. No. 2.6e-22;		
Matches	93;	Conservative	57; Mismatches 152;	Indels 3; Gaps 1;

[illegible]

QY 191 QPEDVNLRAVISEYQKLPDIPIGYSGHETGIALSVAALGAKVLERHITLDTKWGSDH 250
 DB 189 PYEDVRLGMDNLSAEPDAIIGLSDHLDNVACIAGVALGSGIILERHFTDRMORPGDDI 248
 QY 251 SASLEPGEGLAELVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVKIPEGTILTMW 310
 DB 249 VCSMNPDTFKELKQAHALKARGGKDTIAGEXPTKDFAFASVVADKDIKKGELLSD 308
 QY 311 MLTVK 315
 DB 309 NLWVK 313

RESULT 8
 S39722
 spore coat polysaccharide synthesis protein spsE - Bacillus subtilis
 N:Alternate names: protein ipa-66d
 C:Species: Bacillus subtilis
 C>Date: 07-Oct-1994 #sequence_revision 26-May-1995 #text_change 20-Jun-2000
 C:Accession: S39722; C69717
 R:Glaeser, P.; Kuns, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, A.; Rapoport, G.; Danchin, A.
 Mol. Microbiol. 10, 371-384, 1993
 A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr
 A:Reference number: S39655; MUID:95020537; PMID:7934828
 A:Accession: S39722
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-373 <GLA>
 A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51623.1; PID:g413991
 A>Note: the nucleotide sequence was submitted to the EMBL data library, June 1993
 R:Kumet, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet
 C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyana,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: C69717
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-373 <KUN>
 A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15813.1; PID:g2636322
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: spsE
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 19.6%; Score 370; DB 2; Length 373;
 Best Local Similarity 30.1%; Pred. No. 2.7e-21;
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;

QY 13 VGGQHPCTIATIGONHGDLDVAKEMIRMAKCGADCAKFKSELEKFNKALERPPT 72
 DB 11 VGRDAPVFTIAGINHDGLDQAFALIDAAEAGADAVKFM----FQADRMVQKDPGL 66
 QY 73 SKHSGKTYGHEK--RHLEFSDHQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFF 130
 DB 67 YKTAAGKDVSYSLVQSMEMPAEWILPLLDYCREKQVIFLSTVCDGSDALLQSTSPSAF 126
 QY 131 KVGSGDTNPPYLEKTAAGKGRPMVSSGMSQMDTMKQVQIVKPL--NPNFCFLOCTSAVP 189
 DB 127 KIASEYNHLLPLKYYARLNRPMIFSTAGAEISDVHEAWRTIRAEQGNQIAIMHCYAKYP 196

QY 130 LQPEDVNLRAVISEYQKLPDIPIGYSGHETGIALSVAALGAKVLERHITLDTKWGS 248
 DB 187 APPEYSNLSVPLMLAAAPPEAVIGFSDHSEHPTEAPCAVRLGAKLEKHFIDKNLPCA 246
 QY 249 DHSASLEPGEGLAELVRSVRLVERALGSPTKQLLPCMACNEKLGKSV 295
 DB 247 DHSFALNPDELKEMVDGIRKTEAEKLGITKPVSEKLLGSSYKTTTAEIGEIRNFAYRGI 306
 QY 296 VAKVKIPEGTILTMDLTV--KVGE--PKAYPPEDIENLV--GKKVLVTVEEDDTIM 346
 DB 307 FTTAPIQGEAFSEDNIAVLRPGQKQGLHPR--FFELLTSGVRAVRDIPADTGIV 360

RESULT 9
 DB7604
 neuB protein, probable [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: DB7604
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.B.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: DB7604
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-356 <STO>
 A:Cross-references: GB:AE005673; NID:g13424484; PIDN:AAK24832.1; GSPDB:GND00148
 C:Genetics:
 A:Gene: CC2868
 C:Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 19.0%; Score 358; DB 2; Length 356;
 Best Local Similarity 30.1%; Pred. No. 2.3e-20;
 Matches 102; Conservative 55; Mismatches 170; Indels 12; Gaps 9;

QY 1 MPLELELCFGRVWGQHCFFIATIGONHGDLDVAKEMIRMAKCGADCAKFKSELEF 60
 DB 8 IFFSIEIA--GRKIGADHSPYICELSGNHGSLERCLAMVDAADATGCCDAIKIQTITAD- 65
 QY 61 KENKALERPPTYSKHS--W-GKTYGEHKKRHLFSDHQYRELQRYAEVGIFFTAGMDRM 117
 DB 66 --TITLDVDEPEKIHGGLWDGRTLYEYEAHTPEWHAATFERARQGRQVTIFFSFPDET 124
 QY 118 AVEFLHNLNVPFFKVGSGDTNPPYLEKTAAGKGRPMVSSGMSQMDTMKQVQIVKPLN- 176
 DB 125 AVDLLDSLGAPEAKFIASFEAVDPLIKYAAAKGKPLIISTGMANLTEMQTALDTSLSAGA 184
 QY 177 PNFCELOCTSAVPELOPEDVNLRAVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLE 236
 DB 185 PGVLLHLCVSSYPATFADANVRTPDMAARF--GCPIGLSDHTPGTAASVAASVLSGACAVE 243
 QY 237 RHITLDTKWGSDHSGASLEPGEGLAELVRSVRLVERALGSPTKQLLPCMACNEKLGKSV 296
 DB 244 KHFTLARADGGDDAFAFSLPEAFKALVDDTKNAWAALGRAHYDVLGSE--ATSILPERSLY 302
 QY 297 AKVKIPEGTILT--MDMLTVKVGEPKAYPPEDIENLVGKK 334
 DB 303 VTADVKAAGEPLTRANVRSVRPG--NGLPPADLDKVLAKG 339

RESULT 10
 B81275
 N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cyl1317 [imported] - Campylobacter jejuni
 C:Species: Campylobacter jejuni
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
 C:Accession: B81275
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrer
 Nature 403, 665-668, 2000

A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: B81275
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-343 <PAR>
A;Cross-references: GB:AL139078; GB:AL111168; NID:G6968723; PIDN:CAB73744.1; PID:G696875
A;Experimental source: serotype O2, strain NCTC 11168
A;Genetics:
A;Gene: neuB3; Cj1317
C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
C;Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.9%; Score 356.5; DB 2; Length 343;
Best Local Similarity 29.9%; Pred. No. 2.8e-20;
Matches 104; Conservative 65; Mismatches 138; Indels 41; Gaps 10;

Qy 20 FIIAEIGNHQGDLDVAKRMIRMAKCEGADCAKFO-----KSELEFK---PNR 64
Db 13 FIIAEISANHAGSLEMAKLSIKAAKAGADAIAKIQTYTPDSLTLSNDSKEDFIIGLWDK 72

Qy 65 KALERPPTS---KHSWGKTYGEHKKHLEFSDHQVRELQRYABEVGIFFTASGMDMAVEF 121
Db 73 KLYELYESAKTPYEW-----HSQIFET---AQNEGILCFSSPPFAKEDVEF 115

Qy 122 LHELNVPFKVGSGDTNNFPYLEKTAKGRPMWISSGQSDMTWKQVQIVK-PLNPNFC 180
Db 116 LKRPDPTAYKIASPEANDENFVRIAEKKEPTVSTGIATEEELFKICEIPKEEKNPDV 175

Qy 181 FLOCTSAYPELOPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALCAKVLERHIT 240
Db 176 FLKCTSTPYAIEDMNLKGVLSKEF-NVEVGLSDHSFGFLAPMAVALGARVIEKHF 234

Qy 241 LDKTWGSDSHASLEPEGLAEVRSVLVERALGSPTKQLLPCMACNEKLGKSVARVK 300
Db 235 LDKSIESEDSKFSDFDEFKAMVDVAVQAESALGD-GKLDLDEKVLKNRVFARSLYASKD 293

Qy 301 IPEGTITLMDMLTVKVGEPK-AYPEPDIINLVGKVLVTVEEDTIME 347
Db 294 IKKGEMFSEE--NVKSVRPSFGLHPKPYQELLGKKASKDKIKFGDALIKQ 339

RESULT 11
H81318
N-acetylneuraminic acid synthetase (EC 4.1.3.-) Cj1141 [imported] - *Campylobacter jejuni*
C;Species: *Campylobacter jejuni*
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: H81318
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: H81318
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-343 <PAR>
A;Cross-references: GB:AL139077; GB:AL111168; NID:G6968444; PIDN:CAB73396.1; PID:G696857
A;Experimental source: serotype O2, strain NCTC 11168
A;Genetics:
A;Gene: neuB1; Cj1141
C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps
C;Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 18.8%; Score 354; DB 2; Length 343;
Best Local Similarity 30.5%; Pred. No. 4.4e-20;
Matches 96; Conservative 57; Mismatches 156; Indels 6; Gaps 3;

Qy 21 IIAEIGNHQGDLDVAKRMIRMAKCEGADCAKFOKSELEFKFNKALE-RPYTSKHSWGK 79
Db 17 IIPETGINHNGSLEIAKMAKRAKAKIHKQTHIVDEMSQEAKNVPGNANIS--- 73

Qy 80 TYGEHKKHLEFSDHQVRELQRYABEVGIFFTASGMDMAVEFHLNVPFFKVGSGDTNN 139

Db 74 -IYEIMEQCALNYKDELALKVEKQGLVYLSTPFSSRAAANRLEDMGVSAVYIGSGECNN 132
Qy 140 FPYLEKTAKGRPMWISSGQSDMTWKQVQIVKPLNPNFCFLOCTSAYPELOPEDVNLRV 199
Db 133 YPLIKHIAQFKKPMIISGWNISISIKPTVKILRDYEIPFVLLHTTNLYPTPSHLVRLOA 192
Qy 200 ISEYOKLPDIPIGYSGHETGIAISVAALCAKVLERHITLDTKWTGSDSHASLEPEGL 259
Db 193 MLEYLKEP-NCLYGLSDHTTNLACIGAIALGASVLERHFTDTMDRKGPDI VCSMDESTL 251
Qy 260 AELVRSVLVERALGSPTKQLLPCMACNEKLGKSVAKVPIPEGTITLMDMLTVKVGEP 319
Db 252 KDLINQTOEMVLLRGDNNKNPLKEEQVTIDFAFASVWSIKDKIKKGEILSMONIVWRPSK 311
Qy 320 KAYPEPDIINLVGKK 334
Db 312 GGISAKOFEALIGKR 326

RESULT 12
B64542
spore coat polysaccharide biosynthesis protein E - *Helicobacter pylori* (strain 26695)
C;Species: *Helicobacter pylori*
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 20-Jun-2000
C;Accession: B64542
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.N
A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: B64542
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-340 <TON>
A;Cross-references: GB:AB000538; GB:AB000511; NID:G2313263; PIDN:AAD07248.1; PID:G2313267
C;Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 16.1%; Score 303; DB 2; Length 340;
Best Local Similarity 27.7%; Pred. No. 4.3e-16;
Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

Qy 18 PCFIIEIGNHQGDLDVAKRMIRMAKCEGADCAKFOK-----SELEFKFNKALERPPT 72
Db 4 PPKIVAEISANHQDLNLAKESLHAIKESGADFVKLQTYTPSCWTLNSKEDPFIIQGTW 63

Qy 73 SKHSWGKTYGEHKKHLEFSDHQVRELQRYABEVGIFFTASGMDMAVEFHLNVPFFKV 132
Db 64 DKNLYELYOKASTPLEW-HAELFELARKL-DLGIF--SSPFSSQALELLESNCMPMYKI 119

Qy 133 GSGDTNNFPYLEKTAKGRPMWISSGQSDMTWKQVQIVKPLNPNF--CFLOCTSAYPEL 190
Db 120 ASPEIVDLDLIEKAARTQKPIIISGGIATHTTEQDAISLCRRVN-NFDITLLKCVSAYS 178

Qy 191 QPEDVNLRVISEYOKLPDIPIGYSGHETGIAISVAALCAKVLERHITLDTKWTGSDH 250
Db 179 KIEDANLLSMVKLGEIF-GVKFGLSGHTIGSLCPILLATTIGASWIEKHFLNLSLQTPDS 237

Qy 251 SASLEPEGLAEVRSVLVERALGSPTKQLLPCMACNEKLGKSVAKVPIPEGTITL-- 308
Db 238 AFSDMDFNGFKSMVEAIKQSVLALGEEPRINPKTLKRRRFFARSLFVKIDKIQKEALTEN 297

Qy 309 -MDMLTVKVG-EPKAYPEPDIINLVGKK 334
Db 298 NIKALRPNLGLHPEFYK-----EILGQK 320

RESULT 13
B71965
sialic acid synthase - *Helicobacter pylori* (strain J99)
C;Species: *Helicobacter pylori*

A: Variety: strain J99
C: Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Jun-2000
C: Accession: B71965
R: Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A: Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A: Reference number: A71800; MUID: 99120557; PMID: 9923682
A: Accession: B71965
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-340 <ARN>
A: Cross-references: GB:AE001455; GB:AE001439; NID: g4154678; PID: g415468
A: Experimental source: strain J99
C: Genetics:
A: Gene: neuB
C: Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 15.2%; Score 287; DB 2; Length 340;
Best Local Similarity 27.8%; Pred. No. 7.8e-15;
Matches 92; Conservative 68; Mismatches 137; Indels 34; Gaps 14;

QY 21 IIAEIQNHQHGDLVAKRMIRMAKCGADCAKFKQ-----SELEPKFNKALERPYSKH 75
DB 7 IVAELSANHNQDLNLAKESHAIKESGADFVKLTQVTPSCMTLDSK-----EDPFIIQ 60

QY 76 S-WGK--TYGEHR---HLEFSDHQYRELQRYAEVGIFFTAGMDENAVELHNLVPP 129
DB 61 TLWDKENLYGLYQKASTPLEW-HAELELAKKL-DLIGIF--SSPFSSKALELLESLDCPM 116

QY 130 FKVGSGDTNNFPYLEKTAKKGRPMVSSQMGMDTMKVQYQVVKPLNPNF--CFLOQTS 187
DB 117 YIAEFAEIVDLLEKAAKTQPIILSSGIATHAELODAISLCRGW-NFDITLKCVA 175

QY 188 YFLOPEDNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLQTKWG 247
DB 176 YPSKIEDAHLMSVKLGTF-GVKFGLSDHTIGSLCPILATILGASMIKHEFLNKLQT 234

QY 248 SHSASLEGEAELELVRSLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTIL 307
DB 235 PDSAFSMDPNFGKSMVGAIKQSVLALGEEPKINPKTLERRFFARSLFVIKDIQGEAL 294

QY 308 TMD---MLTVKVG-EPKAYPPEDIFNLVGKK 334
DB 295 TSDNIKALPNLGLHKKFYK-----EILGQK 320

RESULT 14
H71307
probable spore coat polysaccharide biosynthesis protein (spse) - syphilis spirochete
C: Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C: Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 19-Apr-2002
C: Accession: H71307
R: Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A: Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A: Reference number: A71250; MUID: 98332770; PMID: 9665876
A: Accession: H71307
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-378 <COL>
A: Cross-references: GB:AE001232; GB:AE000520; NID: g3322856; PID: AAC65539.1; PID: g332285
A: Experimental source: strain Nichols
C: Genetics:
A: Gene: TP0562
C: Superfamily: Streptococcus agalactiae capsular polysaccharide biosynthesis protein cps

Query Match 12.0%; Score 226; DB 2; Length 378;
Best Local Similarity 25.1%; Pred. No. 5.4e-10;
Matches 85; Conservative 47; Mismatches 166; Indels 40; Gaps 9;

QY 8 CPGRWVGGQHPCFIIEIAGIQNHQHGDLVAKRMIRMAKCGADCAKFKQSELEFKFNKAL 67
DB 4 CGGRCFRDPADILITIAEIGSAHAGSDFRABALIDAAADAAAANKFQ-----LIVAHEI 57

QY 68 ERPYTSK---HSGWKTYGEHRKHLEFSDHQYRELQRYAEVGIFFTAGMDENAVELH 124
DB 58 LHPLTGAVRLPSGAVSLYQRFEELVPLSFYAQCENHARSRLVGLISPFGRSATEALA 117

QY 125 LNVPEFKVSGDGTNNFPYLEKT-AKKGRPMVSSGM---OSMDTMKVQYQVVKPLNPNF 179
DB 118 LKPDFLKVASPEL-NYPTLITSLAAEPLILSSGVCLLKEIEGALAQRQYTKQ-GSSH 175

QY 180 CFLQCTSAYPELQPEVDNLRVISEYQKLPDIPIGYSGHETG-IAISVAALGAKVLERH 238
DB 176 ALLHCITAYPAPETEVNLAALLPALATIF-NINVGVDHSVDPLVPLLARAHGACIVEKH 234

QY 239 ITLDTWKSGSHSASLEGEAELELVRSLVRSV-----RLVERALGSP 276
DB 235 ICLSRDTAGLDSDIALDPADFTMTAALNSCARRSPSQIISFLHERGYAPHVVRAVIGSG 294

QY 277 TKQLLPCMACNEKLGKSVVAKVPIEGTILTMDLTV 314
DB 295 EKVLPASRAHYQKNSRLSHYLHAYPRGTIVLOKENLVI 332

RESULT 15
B90173
hypothetical protein aroG [imported] - Sulfolobus solfataricus
C: Species: Sulfolobus solfataricus
C: Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 18-Nov-2002
C: Accession: B90173
R: She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A: Description: Sulfolobus solfataricus complete genome.
A: Reference number: A99139
A: Accession: B90173
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-331 <KUR>
A: Cross-references: GB:AE006641; NID: g13813447; PID: AAK40641.1; GSPDB: GN00155
C: Genetics:
A: Gene: aroG
C: Superfamily: phospho-2-dehydro-3-deoxyoctonate aldolase

Query Match 6.5%; Score 122.5; DB 2; Length 331;
Best Local Similarity 23.9%; Pred. No. 0.059;
Matches 66; Conservative 46; Mismatches 103; Indels 61; Gaps 15;

QY 18 PCFIIEIAGIQNHQHGDLVAKRMIRMAKCGADCAKFKQSELEFKFNKALERPYSKH 77
DB 95 PCAV-----ENEQVUTTKAV-----KRAGASLL-----RGKAYKRTSPYSF 133

QY 78 GKTGYGSHRKHLEFSDHQYRELQRYAEVGIFFTAGMDENAVELHNLVPEFKVSGDT 137
DB 134 -QCLGE-----EGVKILRRVGEVGLPIVTEIMDTRDSNIFSQY-VDMTQICARNA 182

QY 138 NNFPYLEKTAKKGRPMVSSQMGMDTMKVQYQVVKPL-----NENFC-----FLQCTS 186
DB 183 QNFSLLKEVGLKPKVLLKRGMG--NTVEEWLQAEYILLEGNGNTVLCERGIRTEKST 240

QY 187 AVPLQPEDNLRVISEYQKLPDIPIT-GYSGHETG-----IAISVAALGAKVL---ERH 238
DB 241 RFTL---DIGMWAA---KLMTHLPICADPSHPAGKRELVHSLAALAAAGADMILLIEVH 294

QY 239 ITLDTWKSGSHSASLEGEAELELVRSLVRSVRLVERALG 274
DB 295 PHPEKAL--SDSEBQQLTPESFEVLMNRIRTLARALG 328

Search completed: September 13, 2004, 14:44:25
Job time : 48 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:24:08 ; Search time 25 Seconds
(without alignments)
747.727 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MFLELECPGRWVGQHPFCF.....EEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1883	99.8	359	1 SIAS_HUMAN	Qnr45 homo sapien
2	619	32.8	337	1 YA65_METJA	Q58465 methanococ
3	370	19.6	373	1 SPSE_BACSU	P39625 bacillus su
4	113.5	6.0	1248	1 SM1B_MOUSE	Q920f6 mus musculu
5	110	5.8	134	1 ANP3_RHIDE	P35753 rhigophila
6	108.5	5.7	633	1 PARE_MYCGE	P47445 mycoplasma
7	108	5.7	63	1 ANP2_AUSBR	P12101 austrolycic
8	106.5	5.6	478	1 AROF_THEMA	Q9wyh8 thermotoga
9	102.5	5.4	1140	1 GATB_AQUAE	O66766 aquifex aeo
10	102.5	5.4	63	1 YD76_MYCPN	P75405 mycoplasma
11	102	5.4	63	1 ANP1_AUSBR	P12100 austrolycic
12	101	5.4	1107	1 MY1E_RAT	Q63356 rattus norv
13	100.5	5.3	469	1 EGLI_BACSU	P40740 bacillus su
14	100	5.3	1109	1 MY1E_HUMAN	Q12965 homo sapien
15	98.5	5.2	371	1 GPR_BACME	P23231 bacillus me
16	97.5	5.2	642	1 DNAK_XANCP	Q8pak9 xanthomonas
17	97	5.1	378	1 ARGE_VIBPA	P59601 vibrio para
18	96.5	5.1	641	1 DNAK_XANAC	Q8pm00 xanthomonas
19	96.5	5.1	1437	1 DPO3_LACPL	Q88vk2 lactobacill
20	96	5.1	880	1 PODK_RICPR	Q9zd55 rickettsia
21	95.5	5.1	513	1 CHLB_MARPO	P26238 marchantia
22	95.5	5.1	2869	1 RBPI_PLAVB	Q00798 plasmodium
23	94	5.0	64	1 ANP2_RHIDE	P12102 rhigophila
24	93.5	5.0	512	1 GPMI_CLOPE	Q8xku2 clostridium
25	93	4.9	84	1 ANP1_RHIDE	P12015 rhigophila
26	93	4.9	542	1 CYMO_ACTIP	P12015 acinetobact
27	92.5	4.9	1633	1 YP74_CAEEL	Q09221 caenorhabdi
28	92	4.9	346	1 BPHI_BURCE	P51015 burkholderi
29	92	4.9	638	1 DNAK_XYLF	Q9pb05 xylella fas
30	92	4.9	1656	1 ATCC_YEAST	Q12674 saccharomyc
31	91.5	4.8	458	1 EF11_RHIRA	P06805 rhizomucor
32	91.5	4.8	458	1 EF12_RHIRA	P14864 rhizomucor
33	91	4.8	91	1 ANP3_MACAM	P19606 macrozoarce

RESULT 1
SIAS_HUMAN

ID	SIAS_HUMAN	STANDARD;	PRT;	359 AA.
AC	Q9NR45; Q8WUV9; Q9BWS6; Q9NVD4;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Sialic acid synthase (N-acetylneuraminase synthase) (EC 2.5.1.56) (N-acetylneuraminic acid synthase) (N-acetylneuraminic acid phosphate synthase) (EC 2.5.1.57) (N-acetylneuraminic acid phosphate synthase).			
GN	SAS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCHI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RC	TISSUE=Liver;			
RX	MEDLINE=20298869; PubMed=10749855;			
RA	Lawrence S.M., Huddleston K.A., Pitts L.R., Nguyen N., Lee Y.C., Vann W.F., Coleman T.A., Betenbaugh M.J.;			
RT	"Cloning and expression of the human N-acetylneuraminic acid phosphate synthase gene with 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid biosynthetic ability."			
RT	J. Biol. Chem. 275:17869-17877(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y., Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;			
RT	"NEDO human cDNA sequencing project."			
RT	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A., AND VARIANT ASP-68.			
RC	TISSUE=Lung, and Placenta;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieff F., Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L., Shapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Guimond J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."			

ALIGNMENTS

34	91	4.8	197	1	CLPP_BUCAP
35	90	4.8	538	1	CP18_DROME
36	89.5	4.7	381	1	NCAP_CVCAI
37	89.5	4.7	397	1	LE11_METKA
38	89.5	4.7	593	1	MDLB_ECOLI
39	89.5	4.7	638	1	DNAK_XYLF
40	89	4.7	838	1	PTF1_XANCP
41	88.5	4.7	587	1	YNV8_YEAST
42	88	4.7	66	1	ANP1_LYCP
43	88	4.7	479	1	SCRB_STRMU
44	88	4.7	508	1	CHLB_SYNP7
45	88	4.7	855	1	ORP3_MOUSE

Q8k990	buchnera ap
Q95078	drosophila
P36298	canine ente
O8t428	methanopyru
P75706	escherichia
Q87bs8	xylella fas
P45597	x multiphos
P40151	saccharomyc
P24028	lycodes pol
P13522	streptococc
Q8gjn0	synecococc
Q9dbs9	mus musculu

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Produces N-acetylneuraminic acid (Neu5Ac) and 2-keto-3-
CC deoxy-D-glycero-D-galacto-nononic acid (KDN). Can also use N-
CC acetylmannosamine 6-phosphate and mannose 6-phosphate as
CC substrates to generate phosphorylated forms of Neu5Ac and KDN,
CC respectively.
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine +
CC H(2)O = phosphate + N-acetylneuraminic acid.
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + N-acetyl-D-mannosamine 6-
CC phosphate + H(2)O = N-acetylneuraminic acid 9-phosphate + phosphate.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF257466; BAA75261.1; -
DR EMBL; AK001659; BAA91818.1; -
DR EMBL; BC000008; BAA00008.1; -
DR EMBL; BC019315; BAA19315.1; -
DR MIM; 605202; -
DR HSP; P19614; 9AME.
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0008781; F:N-acetylneuraminic acid biosynthesis; NAS.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; NAS.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR004144; NeuB.
DR Pfam; PF01354; Antifreeze; 1.
DR ProDom; PD003258; AntifreezeIII; 1.
DR PROSITE; PS0844; AFP-LIKE; 1.
KW Transferrase; Polymorphism.
FT DOMAIN 294 353 AFP-LIKE.
FT VARIANT 68 68 E -> D.
FT FTID; VAR 013308.
FT CONFLICT 232 232 A -> T (IN REF. 2).
FT CONFLICT 321 321 G -> A (IN REF. 1).
SQ SEQUENCE 359 AA; 40307 MW; 2E02D47F4F98592F CRC64;

Query Match 99.8%; Score 1883; DB 1; Length 359;
Best Local Similarity 99.7%; Pred. No. 7.2e-136;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPLELELCPRWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCGADCAKFKQSELEF 60
DB 1 MPLELELCPRWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCGADCAKFKQSELEF 60

QY 61 KFNKALERYTTSKSHGKTYGHRHLEFSHQYRELQRYAEVGIFFFTASGMDMAVE 120
DB 61 KFNKALERYTTSKSHGKTYGHRHLEFSHQYRELQRYAEVGIFFFTASGMDMAVE 120

QY 121 FLHELNVPPFKVSGDGNFNPYLEKTAKKGRPMVSISSGMSMDTMKVQYIVKPLNPF 180
DB 121 FLHELNVPPFKVSGDGNFNPYLEKTAKKGRPMVSISSGMSMDTMKVQYIVKPLNPF 180

QY 181 FLOCTSAIPQEDVNLRVISYOKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240
DB 181 FLOCTSAIPQEDVNLRVISYOKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240

QY 241 LDKTWKSDRSASLEPCELAEVLRSVRLVERALGSPTKQLLPCEMAKNEKLGKSVVAKV 300
DB 241 LDKTWKSDRSASLEPCELAEVLRSVRLVERALGSPTKQLLPCEMAKNEKLGKSVVAKV 300

QY 301 IPEGTILTMDMLTVKVGPKAPYPEDIPNLVGGKVLVTVEBDDTIMBELVDNHGKTKS 359
DB 301 IPEGTILTMDMLTVKVGPKAPYPEDIPNLVGGKVLVTVEBDDTIMBELVDNHGKTKS 359

```

```

RESULT 2
YAG5_METJA STANDARD; PRT; 337 AA.
ID YAG5_METJA
AC Q58465;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-JUL-1998 (Rel. 36, Last sequence update)
DT 28-JUL-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein MJ1065.
GN MJ1065.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcales; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadov P.W., Hanna M.C.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii";
RL Science 273:1058-1073 (1996).
CC -!- SIMILARITY: STRONG, TO B.SUBTILIS SPSE.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67549; AAB99068.1; -
DR PIR; H64432; H64432.
DR TIGR; MJ1065; -
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; NeuB.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; NeuB; 1.
DR PROSITE; PS0844; AFP-LIKE; 1.
KW Hypothetical protein; Complete proteome.
FT DOMAIN 279 337 AFP-LIKE.
FT SEQUENCE 337 AA; 37976 MW; 790CAFF48C111B CRC64;

Query Match 32.8%; Score 619; DB 1; Length 337;
Best Local Similarity 40.0%; Pred. No. 8.1e-40;
Matches 136; Conservative 65; Mismatches 115; Indels 24; Gaps 7;

QY 11 RWVGQHPGCFIIAIGONHOGDLDVAKRMIRMAKCGADCAKFKQSELEFNRKALERP 70
DB 9 RYVGKGEPTIIAEGGLNHGDDIGKELVKEAKCGADAIKFSYHTE-----D 58

QY 71 YTSKSHGKTYGHRHLEFSHQYRELQRYAEVGIFFFTASGMDMAVEFLHELNVPPF 130
DB 59 FISKKS--EYELFKSLSESEFEVELKEYAEKIGIMFISTPLDLKYVDILNKNVPF 115

QY 131 KVGSGDGNFNPYLEKTAKKGRPMVSISSGMSMDTMKVQYIVKPLNPF 186
DB 116 KIASGDLTFFPLEKVAKTGKPVILSTGMSDIG---EIVEAVKVLNNGCDRIILHCIS 172

QY 187 AYLQPEDVNLRVISYOKLPDPIPIGYSGHETGIAISVAALGAKVLERHITLDKTKW 246
DB 173 SYTPYEDVNLNALKTKLSIF-NIPVGYSDHTLGLILAPVVSVALGADVIEKHTLDKME 231

```

QY 247 GSDHSALEPGLAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVAKVKIPEGTI 306
 DB 232 GPDHALSADPEFEKEMWNNIRLVERMLGSGEKIPWSESDVIVARRSIVAKRNKKGEY 291
 QY 307 LTMDLTKVKGEP-KAYPEPDIENFVNGKKLVTVVEEDDTI 345
 DB 292 LSVNDNISFK-RPGRGIETKVLISILNRKIKNKDEEDDII 329

RESULT 3
 ID SPSE BACSU STANDARD; PRT; 373 AA.
 AC P39625;
 DT 01-FEB-1995 (Rel. 31, Last Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Spore coat polysaccharide biosynthesis protein spsE.
 GN SPSE OR IPA-67D OR BSU37870.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serrero P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 CC -!- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -!- SIMILARITY: Contains 1 AFP-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; X73124; CAA51623.1; -;
 DR EMBL; Z99123; CABI5813.1; -;
 DR FIR; S39722; S39722.
 DR Subtilist; BG10613; spsE.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze_like.
 DR InterPro; IPR004144; Neut.
 DR Pfam; PF01354; Antifreeze; 1.
 DR Pfam; PF03102; Neut; 1.
 DR PROSITE; PS0844; AFP_LIKE; 1.
 KW Complete proteome.
 FT DOMAIN 305 367 AFP-LIKE.
 SQ SEQUENCE 373 AA; 40889 MW; 3CFBDF6ACE0DBCE8 CRC64;
 Query Match 19.6%; Score 370; DB 1; Length 373;
 Best Local Similarity 30.1%; Pred. No. 7.8e-21;
 Matches 107; Conservative 55; Mismatches 167; Indels 26; Gaps 9;
 QY 13 VGGHPCFIITAEIQNHQGLDVAKRMIRMAKEGACAKPKSELEFKNRKALERPVT 72
 DB 11 VGKDAPVFIIAEGINHDGLDQAFALIDAAEAGADAVKFM----FQADRMVYQKDPGL 66
 QY 73 SKHSWGTYGHEK--RHLEFSDHQYRELQRYAERVGIFFTASGMDENAVELHNLNPPFF 130
 DB 67 YKTAAGKDVIFSLVQSMENPAEWILPDLYCREKQVIFLSTVDCGSDADLQSTSPSAF 126
 QY 131 KVGSGDTNPNFYLEKTAAGKGRPMVVISGMSQMDTMKVQIVKPL-NPNFCFLQCTSAYP 189
 DB 127 KIASYEINHLPLKYYVARLNRPMTFTAGAEISDVHEAWRTIRAEQNNQTAIMHCVAKYP 186
 QY 190 LQEDVNLRVISEYKQLPDPIDIGYSGH-ETGIAISVAALGAKVLERHITLDKTKWGS 248
 DB 187 APPEYNSLVIPMLAAAPPEAVIGFSDHSBHTPEAPCAAVRLGAKLIEKHTIDKNLPGA 246
 QY 249 DHSASLEPGLAELVRSV-----LVERALGSPTKQLLPCEMACNEKLGKSV 295
 DB 247 DHSFALNPDELKEMVDIRTEALQKGTVPSEKLLGSSYKTTTAEIGEIRNFAYRGI 306
 QY 296 VAKVKIPEGTILTMDLTV-KVGE-KRAYPEPDIENFV--GKKVLTVVEEDDTIM 346
 DB 307 FTTAPIQKGEAFSEDNIAVLRLPGKQKQLHPR-FFELLTSGVRAVRDIPADTGIV 360

RESULT 4
 ID SM1B MOUSE STANDARD; PRT; 1248 AA.
 AC Q920F6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosomes 1-like 2 protein (SMC1beta
 DE protein).
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A., PARTIAL SEQUENCE, FUNCTION, CHARACTERIZATION, AND
 RP INTERACTION WITH SMC3.
 RC TISSUE-Testis;
 RX MEDLINE=21448993; PubMed=11564881;
 RA Revenkova E., Eijpe M., Heyting C., Gross B., Jessberger R.;
 RT "Novel meiosis-specific isoform of mammalian SMC1.";
 RL Mol. Cell. Biol. 21:6984-6998(2001).
 CC -!- FUNCTION: Meiosis specific component of cohesin complex. The
 CC cohesin complex is required for the cohesion of sister chromatids
 CC after DNA replication. The cohesin complex apparently forms a

large proteinaceous ring within which sister chromatids can be trapped. At anaphase, the complex is cleaved and dissociates from chromatin, allowing sister chromatids to segregate. The meiosis-specific cohesin complex probably replaces mitosis specific cohesin complex when it dissociates from chromatin during prophase I.

-!- SUBUNIT: Forms a heterodimer with SMC3. Component of a meiosis-specific cohesin complex, probably composed of the SMC2L2 and SMC3 heterodimer attached via their hinge domain, RAD21 (or its meiosis-specific related protein REC8), which link them, and STAG3, which interacts with RAD21 or REC8.

-!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin. In prophase I stage of meiosis, it is found along the axial elements of synaptonemal complexes. In late-pachytene-diplotene, the bulk of protein dissociates from the chromosome arms probably because of phosphorylation by PLK, except at centromeres, where cohesin complexes remain. It however remains chromatin associated at the centromeres up to metaphase II. At anaphase II, it dissociates from centromeres, allowing chromosomes segregation.

-!- TISSUE SPECIFICITY: Testis and ovary specific. Not expressed in somatic cells.

-!- DOMAIN: The flexible hinge domain, which separates the large intramolecular coiled coil regions, allows the heterotypic interaction with the corresponding domain of SMC3, forming a V-shaped heterodimer. The two heads of the heterodimer are then connected by different ends of the cleavable RAD21 or REC8 protein, forming a ring structure (By similarity).

-!- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF033827; AAL09333.1; -
 MGD; MGI:2154049; Smc112.
 GO; GO:0000775; C:chromosome, pericentric region; IDA.
 GO; GO:0000795; C:synaptonemal complex; IDA.
 GO; GO:0003677; F:DNA binding; IDA.
 GO; GO:0005515; F:protein binding; IPI.
 GO; GO:0007062; P:sister chromatid cohesion; IDA.
 InterPro; IPR003439; ABC transporter.
 InterPro; IPR003405; SMC C.
 InterPro; IPR003395; SMC N.
 Pfam; PF02483; SMC C; 1.
 Pfam; PF02463; SMC N; 1.
 Meiosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil; Nuclear protein.
 NP BIND 32 39 ATP (POTENTIAL).
 FT DOMAIN 163 502 COILED COIL (POTENTIAL).
 FT DOMAIN 503 665 FLEXIBLE HINGE.
 FT DOMAIN 666 912 COILED COIL (POTENTIAL).
 FT DOMAIN 1123 1158 ALA/ASP-RICH (DA-BOX).
 FT SEQUENCE 1248 AA; 144513 MW; 6C31DB46217BC94D CRC64;
 Query Match 6.0%; Score 113.5; DB 1; Length 1248;
 Best Local Similarity 23.5%; Pred. No. 1.1;
 Matches 88; Conservative 56; Mismatches 138; Indels 93; Gaps 19;
 31 GDLVAKRMIRMAKEGACADCAKEQKSELEKFNK---ALERYPTSGHS-WGTYGEHGR 86
 166 GEYEAKKKLOKAE-----DAQPHFNKKNVAAER-----KHAKIEKEEAHYQ 210
 87 H-----LEFSDHQYRELQRYAEVGIPTASGMDEMAVEFLHNVPPFKVSGDPT-NNFP 141
 211 NLLEELKINKIQLMFLQYNEEKINVLNTELEQM-----DGNLSVVK-----DTLSHHE 260
 142 YLEKTAKKGRPMVSSGMQMDTMKVQYIVKPLNPNFCFLQCTSAVPLOPEDVNLNVIS 201
 261 NIFAKKKDYGMLTRQLQQTAKELKSVEAILNQKRPQVIRAKENTSHHLKLDLSKKLIT 320

202 EYOKLFPDIPGYSGHETGIAISVAVALGAKVLERHITLDTKWGSDHSAS-----LEDGE 258
 321 DNEK-----QCSKQEDGIRALVAELA-----DLDRAWKSPFKQMEKILQKGR 363
 259 LAELIVRS-----VRLVERALGSPTKQLLPC--EMACNEKLG-----KS 294
 364 DIELNSQLDYKLLKEQVRKVGIMTQQLKEQKAEKELAFERKRRHGDTOGNIKQ 423
 295 VVAKV-----KIPGSTILTMQMLTVKVGEPKAYPPEDIFNLVGGKVLTVBEDDTIM 346
 424 IKEQIEEHKKRIEKLLEYTKTQMCLECKKQOEALKE-IENTKSRMSEVN-EELSILIR 481
 347 EEL-----VDNH-GKK 356
 482 NELQAGIDNHEGKR 496

RESULT 5
 ANP3 RHIDE STANDARD; PRT; 134 AA.
 ID ANP3 RHIDE STANDARD; PRT; 134 AA.
 AC P35753.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Antifreeze peptide RD3.
 OS Rhigophila dearborni (Antarctic eel pout) (Lycodichthys dearborni).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidae;
 OC Zoarcidae; Lycodichthys.
 NCBI_TaxID=8201;
 RN [1]
 RP SEQUENCE.
 MEDLINE=95210308; PubMed=7696304;
 RX Wang X., Devries A.L., Cheng C.-H.C.;
 RA "Antifreeze peptide heterogeneity in an antarctic eel pout includes
 RT an unusually large major variant comprised of two 7 kDa type III AFPs
 RT linked in tandem";
 RL Biochim. Biophys. Acta 1247:163-172(1995).
 [2]
 RP STRUCTURE BY NMR OF 1-73.
 RX MEDLINE=99353969; PubMed=10423534;
 RA Miura K., Ohgaya S., Hoshino T., Nemoto N., Odaira M., Nitta K.,
 RA Tsuda S.;
 RT "Determination of the solution structure of the N-domain plus linker
 of antarctic eel pout antifreeze protein RD3.";
 RL J. Biochem. 126:387-394(1999).
 CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point by
 CC absorbing ice and inhibiting its growth.
 CC -!- SIMILARITY: Belongs to the type-III AFP family.
 CC -!- SIMILARITY: Contains 2 AFP-like domains.
 CC PIR; S53514; S53514.
 DR PDB; 3NLA; 23-FEB-99.
 DR PDB; 3RDN; 23-FEB-99.
 DR PDB; 1C89; 28-FEB-01.
 DR PDB; 1C8A; 28-FEB-01.
 DR InterPro; IPR006014; Antifreeze dom.
 DR InterPro; IPR006190; Antifreeze-like.
 DR InterPro; IPR006013; AntifreezeII.
 DR Pfam; PF01354; Antifreeze; 2.
 DR PRINTS; PR00357; ANTIFREEZIII.
 DR ProDom; PD003258; AntifreezeIII; 2.
 DR PROSITE; PS00844; AFP LIKE; 2.
 DR Antifreeze protein; Multigene family; Repeat; 3D-structure.
 KW DOMAIN 4 63 AFP-LIKE 1.
 FT DOMAIN 65 70 LINKER.
 FT DOMAIN 74 133 IMPORTANT FOR ICE-BINDING
 FT SITE 9 9 (BY SIMILARITY).
 FT SITE 14 14 IMPORTANT FOR ICE-BINDING
 FT SITE 18 18 (BY SIMILARITY).
 FT SITE 18 18 IMPORTANT FOR ICE-BINDING

```

CC CC      -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL; U39700; AAC71421.1; -;
CC CC      EMBL; U25549; AAC43990.1; -;
CC CC      PIR; D64222; D64222.
CC CC      HSSP; P06982; IAU6.
CC CC      TIGR; MG203; -.
CC CC      InterPro; IPR003594; ATPbind_ATPase.
CC CC      InterPro; IPR002288; DNA_gyraseB_C.
CC CC      InterPro; IPR001241; DNA_topoisII.
CC CC      InterPro; IPR005740; TopoIV_B_gpos.
CC CC      InterPro; IPR006171; Toprim_dom.
CC CC      Pfam; PF00204; DNA_gyraseB; 1.
CC CC      Pfam; PF02518; HATPase_c; 1.
CC CC      Pfam; PF01751; Toprim_1.
CC CC      PRINTS; PR00418; TP12FAMILY.
CC CC      ProDom; PD149633; DNA_gyraseB_C; 1.
CC CC      SMART; SM00387; HATPase_c; 1.
CC CC      SMART; SM00433; TOP2c; 1.
CC CC      TIGRfams; TIGR01058; parE_gpos; 1.
CC CC      PROSITE; PS00177; TOPOISOMERASE_II; 1.
CC CC      Isomerase; topoisomerase; ATP-binding; Complete proteome.
CC CC      SEQUENCE 633 AA; 71569 MW; 249B854169A88CB6 CRC64;
CC CC
CC CC      Query Match          5.7%; Score 108.5; DB 1; Length 633;
CC CC      Best Local Similarity 18.1%; Pred. No. 1.1;
CC CC      Matches 77; Conservative 63; Mismatches 139; Indels 147; Gaps 18;

```

[illegible]

```

AC P12101;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antifreeze peptide AB2.
OS Austrolycichthys brachycephalus (Antarctic eel pout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidei;
OC Zoarcidae; Austrolycichthys.
OX NCBI_TaxID=8195;
RN [1]
RP SEQUENCE.
RX MEDLINE=89323219; PubMed=2752054;
RA Cheng C.-H.C., Devries A.L.;
RT "Structures of antifreeze peptides from the antarctic eel pout,
RT Austrolycichthys brachycephalus."
RL Biochim. Biophys. Acta 997:55-64(1989).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point by
CC absorbing ice and inhibiting its growth.
CC -!- SIMILARITY: Belongs to the type-III AFP family.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
DR PIR: S04974; S04974.
DR HSP: P35753; 3RD.
DR InterPro: IPR006014; Antifreeze dom.
DR InterPro: IPR006190; Antifreeze_1like.
DR InterPro: IPR006013; AntifreezeII.
DR Pfam: PF01354; Antifreeze; 1.
DR PRINTS: PR00357; ANTIFREEZEIII.
DR ProDom: PD003258; AntifreezeII; 1.
DR PROSITE: PS08444; AFP_LIKE; 1.
DR Antifreeze protein.
KW DOMAIN 3 62 APP-LIKE.
FT SITE 8 8 IMPORTANT FOR ICE-BINDING
FT SITE 13 13 (BY SIMILARITY).
FT SITE 13 13 IMPORTANT FOR ICE-BINDING
FT SITE 17 17 (BY SIMILARITY).
FT SITE 17 17 IMPORTANT FOR ICE-BINDING
FT SITE 43 43 (BY SIMILARITY).
FT SITE 43 43 IMPORTANT FOR ICE-BINDING
FT SITE 43 43 (BY SIMILARITY).
FT SITE 43 43 (BY SIMILARITY).
SQ SEQUENCE 63 AA; 7001 MW; A1E7F954598DD061 CRC64;

Query Match 5.7%; Score 108; DB 1; Length 63;
Best Local Similarity 42.6%; Pred. No. 0.07;
Matches 26; Conservative 10; Mismatches 25; Indels 0; Gaps 0;

QY 293 KSVVAKVPIEGTILTMDLTVKVGPKAYPPEDIFNLGKYLVTVEEDDTIMEELVDN 352
DB 2 KSVVQNQLIPINTALTILVMKAEVSPKGPABEIPRLVGMQVNRVYLDETLMPDWKN 61

QY 353 H 353
DB 62 Y 62

RESULT 8
AROF THEME STANDARD; PRT; 338 AA.
AC Q9WH8;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 2.5.1.54) (Phospho-2-
DE keto-3-deoxyheptonate aldolase) (DAHP synthetase) (3-deoxy-D-arabino-
DE heptulosonate 7-phosphate synthase).
GN AROF OR TM0343.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.W.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima."
RL Nature 399:323-329(1999).
RN [2]
RP CHARACTERIZATION.
RX STRAIN=MSB8 / DSM 3109 / ATCC 43589;
RX MEDLINE=22753848; PubMed=12743122;
RA Wu J., Howe D.L., Woodard R.W.;
RT "Thermotoga maritima 3-deoxy-D-arabino-heptulosonate 7-phosphate
RT (DAHP) synthase: the ancestral eubacterial DAHP synthase?";
RL J. Biol. Chem. 278:27525-27531(2003).
CC -!- FUNCTION: Catalyzes the condensation of phosphoenolpyruvate (PEP)
CC and D-erythrose-4-phosphate (E4P) giving rise to 3-deoxy-D-
CC arabino-heptulosonate-7-phosphate (DAHP).
CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + D-erythrose 4-phosphate
CC + H(2)O = 3-deoxy-D-erythro-hept-2-ulosonate 7-phosphate +
CC phosphate.
CC -!- COFACTOR: Requires divalent ions.
CC -!- ENZYME REGULATION: Inhibited by L-phenylalanine and L-tyrosine.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC first step.
CC -!- SUBUNIT: Homotetramer.
CC -!- MISCELLANEOUS: Optimal pH is 6.3 at 60 degrees Celsius. Maximal
CC activity is at 90 degrees Celsius. Extremely thermostable.
CC -!- SIMILARITY: Belongs to class-I DAHP synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE001715; AAD35429.1; -
CC FIR; E72388; E72388.
CC TIGR; TM0343; -
CC InterPro: IPR006218; DAHP1/KDSA.
CC InterPro: IPR006268; DAHP syn 2.
CC Pfam: PF00793; DAHP_synth_1; 1.
CC TIGRFAMS; TIGR01361; DAHP_synth_Bsub; 1.
KW Aromatic amino acid biosynthesis; Transferase; Complete proteome.
SQ SEQUENCE 338 AA; 37378 MW; E9634B6704D3DF4D CRC64;

Query Match 5.7%; Score 108; DB 1; Length 338;
Best Local Similarity 22.4%; Pred. No. 0.57;
Matches 54; Conservative 50; Mismatches 91; Indels 46; Gaps 10;

QY 56 SELEPKFNKALRBPYTSKHSWKTYGEHKRHFSDHQYRELQRYAEVGVIFETASGMD 115
DB 118 SELGVKVLGGAYKPRTPSPYSF---QGLGKGLLEY-----LREAADKTYGVVVTBALG 167
QY 116 EMAVEFELHNLVFFKVGSGDTNNFVLEKTKGRPMWISSGMQSDMTKWQVQIVKPL 175
DB 168 EDDLPKVAYEY-ADIIQIGARNAQNFRLLSKAGSYNKPVLKRGF--MNTIEFLLSEYI 224
QY 176 --NPNFCFLQC-----TSAYLPQPDVNLRLVISEYOKLFPDIPIGYSG- 216
DB 225 ANSGNTKILICRGIRTFEXATRNLTLDISAVPIRKESHPIL-----VDPSSHSGG 275
QY 217 -HETGIAISVAVALGAK--VLERHIFLDTKWGSDHSASLEPGELAEVRSVLRVREAL 273
DB 276 RRDVLVPLSRAAIVAGHGIIVEVHPPEKAL--SDGKQLDFELFKELVQEMKKIADAL 333
QY 274 G 274

```

```

Db          334 G 334
RESULT 9
GATB_AQUAE STANDARD; PRT; 478 AA.
AC O66766;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl-glutamyl-tRNA(Asn/Gln) amidotransferase subunit B
DE (EC 6.3.5.-) (Asp/Glu-ADT subunit B).
GN GATB OR AQ 461.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VFS;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: Allows the formation of correctly charged Asn-tRNA(Asn)
CC or Gln-tRNA(Gln) through the transamidation of misacylated Asp-
CC tRNA(Asn) or Glu-tRNA(Gln) in organisms which lack either or both
CC of asparaginyl-tRNA or glutamyl-tRNA synthetases. The reaction
CC takes place in the presence of glutamine and ATP through an
CC activated phospho-Asp-tRNA(Asn) or phospho-Glu-tRNA(Gln) (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
CC -!- CATALYTIC ACTIVITY: ATP + L-aspartyl-tRNA(Asn) + L-glutamine = ADP
CC + phosphate + L-asparaginyl-tRNA(Asn) + L-glutamate.
CC -!- SUBUNIT: Heterotrimer of A, B and C subunits (By similarity).
CC -!- SIMILARITY: Belongs to the gatB/gatE family. GatB subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000691; AAC06727.1; -
CC PIR; B70342; B70342.
CC HAMAP; MF_00121; -.
CC InterPro; IPR004413; GatB.
CC InterPro; IPR006107; GatB_cent.
CC InterPro; IPR006075; GatB_N.
CC InterPro; IPR003789; GatB_Yqey.
CC Pfam; PF01162; GatB; 1.
CC Pfam; PF02934; GatB_N; 1.
CC Pfam; PF02637; GatB_Yqey; 1.
CC TIGRFAMs; TIGR00133; gatB; 1.
CC PROSITE; PS01234; GATB; 1.
CC Protein biosynthesis; Ligase; Complete proteome.
CC SEQUENCE 478 AA; 55040 MW; F725AE78944BD79A CRC64;
Query Match 5.6%; Score 106.5; DB 1; Length 478;
Best Local Similarity 19.6%; Pred. No. 1.1;
Matches 62; Conservative 54; Mismatches 100; Indels 101; Gaps 14;
Qy 93 DQYRELQRYAEVGIFFTAGMDKAVYEFLLHNLVPPFKVSGD-----TNNFPYL 143
Db 171 EKLRLNRYA---GV---SKADMEKGLQRLDINVSIRPKGSKFGTRVEIKNVNSRFV 223
Qy 144 EKTAKGRPMVISGMSMDTMKVQIVKPLNPNFCFLQCTSAVPLQ----- 191

```

```

Db          334 G 334
RESULT 10
YD76 MYCPN STANDARD; PRT; 1140 AA.
AC P75405;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN376 (A19_orf1140).
GN MPN376 OR MP460.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmatataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl B., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC MPN375.
CC -!- SIMILARITY: IN THE CENTRAL SECTION; STRONG, TO M.PNEUMONIAE
CC MEN374.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; STRONG, TO M.PNEUMONIAE
CC MPN373.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000045; AAB96108.1; -
CC PIR; S73786; S73786.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 8 28 POTENTIAL.
CC FT TRANSMEM 1098 1118 POTENTIAL.
CC SQ SEQUENCE 1140 AA; 130383 MW; 8FA9406C57DD8886 CRC64;
Query Match 5.4%; Score 102.5; DB 1; Length 1140;
Best Local Similarity 21.3%; Pred. No. 6.7;
Matches 84; Conservative 60; Mismatches 167; Indels 83; Gaps 16;
Qy 22 IAEIGQHQHGDLDVAKMIRMA-----KEGADCAKFOK--SELEFKFNKALERP 71
Db 598 VSAIFQDEGGLRDASPINKVATPKSSKTKIANGASKIEAKANEISYKYH-----YRQN 653
Qy 72 TSKHGWCKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMD--EMAVEFLHNLVPP 129
Db 654 FKQASWDKQNSQTSKIVVQSTDLNDRERFQKDNINLYKVGISETEIKVNAHKVDA-M 712

```

```
QY 130 FKVGSGDTNNPPYLEKTKAGRPWISSQMSDMMKQVQIVKPLNP-----178
Db 713 LNARKSDPKLASVOSTANK-----YGLNLRSNPTVGQFYVVDVTNANDLGNQRANNA 767
QY 179 ---FCFLQ-----CTSAVLPQEDVNLRLVISEYQKLFDPDIPICYSGHETGIAISVAA---227
Db 768 KSYFYIEGLDKGAQSSYLVRFE--NKQKLSLESIAVDSRGLYKVKVSKDALIOAKQKQ 825
QY 228 -----VALGAKVLERHITL-----DKTWKSGDSHASLEPGELAEIVRSVRL--VE 270
Db 926 NLYLDTHNNAALKAMLTNAELTLPASADNSAKLSTPNAENDEGLSENVSGLIGYVE 885
QY 271 RALG-----SPTKQLPCMACNEKLGKSVAKKIPETCTILTMDM-----LT 313
Db 886 RWTGKKLFLKERVSNFKEDKNLKLRTSNFTLDKKNLEVKDP--SVINQIVBEAKGVN 943
QY 314 KVGGEPKAYPPEDIFNLVGKVLVTVEEDDTIME 347
Db 944 VLVSEKGDPPESDKNIF-KIILTTNPQSTVIK 976
```

```
RESULT 11
ANP1_AUSBR STANDARD; PRT; 63 AA.
AC P12100;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antifreeze peptide AB1.
OS Austrolycichthys brachycephalus (Antarctic eel pout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Zoarcoidei;
OC Zoarcidae; Austrolycichthys.
OX NCBI_TaxID=8195;
RN [1]
RP SEQUENCE.
RX MEDLINE=89323219; PubMed=2752054;
RA Cheng C.-H.C.; Devries A.L.;
RT "Structures of antifreeze peptides from the antarctic eel pout,
  Austrolycichthys brachycephalus."
RL Biochim. Biophys. Acta 997:55-64(1989).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point by
  absorbing ice and inhibiting its growth.
CC -!- SIMILARITY: Belongs to the type-III AFP family.
CC -!- SIMILARITY: Contains 1 AFP-like domain.
DR PIR; S04973; S04973.
DR HSSP; P35753; 3RDN.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR006013; AntifreezeII.
DR Pfam; PF01354; Antifreeze; 1.
DR PRINTS; PR00357; ANTIFREEZII.
DR ProDom; PD003258; AntifreezeII; 1.
DR PROSITE; PS50844; AFP_LIKE; 1.
KW Antifreeze protein..
FT DOMAIN 3 62 AFP-LIKE.
FT SITE 8 8 IMPORTANT FOR ICE-BINDING
  (BY SIMILARITY).
FT SITE 13 13 IMPORTANT FOR ICE-BINDING
  (BY SIMILARITY).
FT SITE 17 17 IMPORTANT FOR ICE-BINDING
  (BY SIMILARITY).
FT SITE 43 43 IMPORTANT FOR ICE-BINDING
  (BY SIMILARITY).
FT SEQUENCE 63 AA; 6846 MW; AC84FD14247193B4 CRC64;
```

Query Match 5.4%; Score 102; DB 1; Length 63;
Best Local Similarity 39.3%; Pred. No. 0.2;
Matches 24; Conservative 10; Mismatches 0; Gaps 0;

QY 293 KSVVAKVKIPEGTILTMDMLTVKVGEKAYPPEDIFNLVGKVLVTVEEDDTIMEBLVDN 352

```
Db 2 KSVVASQLIPINTALTPAMKAKVSPKGIPEAKSVKGMQVNRVAVNLDITLMPDWKVT 61
QY 353 H 353
Db 62 Y 62
```

RESULT 12

```
MYIE_RAT STANDARD; PRT; 1107 AA.
AC Q63356;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myosin Ie (Myosin heavy chain myr 3).
GN MYOIE OR MYR3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley;
RX MEDLINE=95247829; PubMed=7730414;
RA Stoeffler H.E.; Ruppert C.; Reinhard J.; Bahler M.;
RT "A novel mammalian myosin I from rat with an SH3 domain localizes to
  Con A-inducible, F-actin-rich structures at cell-cell contacts."
RL J. Cell Biol. 129:819-830(1995).
CC -!- FUNCTION: Myosins are actin-based motor molecules with ATPase
  activity. Unconventional myosins serve in intracellular movements.
  Their highly divergent tails are presumed to bind to membranous
  compartments, which would be moved relative to actin filaments (By
  similarity).
CC -!- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -!- SIMILARITY: Contains 1 IQ domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74815; CA452815.1; -.
DR PIR; S52517; S52517.
DR HSSP; P08799; 1MND.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
  Multigene family.
FT DOMAIN 1 679 MYOSIN HEAD-LIKE.
FT DOMAIN 695 724 IQ.
FT DOMAIN 1050 1107 SH3.
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 112 119 ATP (POTENTIAL).
FT SEQUENCE 1107 AA; 126826 MW; B9D8FBB0CE047148 CRC64;
```

Query Match 5.4%; Score 101; DB 1; Length 1107;

	Best Local Similarity	22.2%;	Pred. No. 8.4;	
	Matches	70;	Conservative	43; Mismatches 114; Indels 88; Gaps 15;
Qy	47	GADCAKFOKSELEF----	KFNKALRPERPYTSKHSGWTKYGBHKLHFSDHOYRELQRYA	102
Dd	476	GADQTLLQLKMQIGSHHEFN-----	-SWNQFIHHVAGKVSYDMGDFCERNR	523
Qy	103	BEVGIFFTASGMDEMAVEFLHMLNVPFKVGGDTNNPFFYLEKTAKRGPRMVISGMSQM	162	
Dd	524	D-----VLFD-----MD-----LIELMOSSLPTIK-----SLFPENLOAQDKGRPTTAGSKIK--	566	
Qy	163	DTMKOVQIVKPLNPFCFLQCTSAI-----	-PLOPEDVNLRVISEYQKLFDPDIP	211
Dd	567	---KQANDLVSTL-----MKCTPHYRCIKPNETKPKPMEESSRVKHQVEYLGLKENIR	617	
Qy	212	IGYGSHGTGAISVAVALGAKVLERHTIDK-----TWKSGDHSASLEPGELAEVRSVR	267	
Dd	618	VRRAGY-----AYRRVQKFQRLORVAILTKATPWVRGDKGCVL-----HLLOSVM	663	
Qy	268	LVERALGSPTKQLLPCEMACNEKLGKSVAKVIPEGTILTMDLTVKVGEPKAYPPEDI	327	
Dd	664	M-----DSQF-----QLGRSKVF- IKAPESILFLEEEMRKERYDGVARVIQKW	706	
Qy	328	FNLVGKKVLTVEED	342	
Dd	707	RKFVARKKYVQMRD	721	
RESULT 13	BGL1_BACSU	ID	STANDARD;	PRT; 469 AA.
AC	PD40740; O32287;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DE	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Beta-glucosidase (EC 3.2.1.21) (Gentiobiase) (Cellobiase) (Beta-D-			
DE	glucoside glucohydrolase) (Amygdalase).			
GN	BGLH OR N17D OR BSU3260.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			
NCBI_TaxID=1423;	(1)			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=168 / Marburg;			
RC	MEDLINE=95189730; PubMed=7883710;			
RA	Le Coq D., Lindner C., Krueger S., Steinmetz M., Stuelke J.;			
RT	"New beta-glucoside (bgl) genes in Bacillus subtilis: the bglp gene			
RT	product has both transport and regulatory functions similar to those			
RT	of BglF, its Escherichia coli homolog.";			
RN	J. Bacteriol. 177:1527-1535(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168 / BGSCIAL;			
RC	MEDLINE=95219088; PubMed=7704263;			
RA	Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;			
RT	"Cloning and sequencing of a 29 kb region of the Bacillus subtilis			
RT	genome containing the hut and wapA loci."			
RT	Microbiology 141:337-343(1995).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RC	MEDLINE=98044033; PubMed=9384377;			
RA	Kunst F., Ogawara N., Mosser I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchart S.,			
RA	Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,			
RA	Denicot F., Devine K.M., Dutethoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferraci E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holtsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			

RA	Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigic C.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro T., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Pujić P., Purnelle B., Rapoport G., Rev M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Taconin E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambut R., Wedler E., Wedler H., Weitzengesser T., Yata K.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;			
RT	The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis.";			
Nature	390:249-256(1997).			
CC	-!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-			
CC	glucose residues with release of beta-D-glucose.			
CC	-!- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to licenses@isb-sib.ch).			
EMBL:	Z3484287.1; --			
DR	EMBL; D31856; BAA			

321 -----KSGNLLGVXNPYLKSEWGWQIDPKGLRITLNTLYDRYQKP-LFIVENGSL 371

329 NLVGGKVLTVVEEDDTIMBELVDNH 353
372 GAVDK-----VEEDGTIQDDYRINY 391

RESULT 14

MYIE HUMAN
ID MYIE HUMAN STANDARD; PRT; 1109 AA.
AC Q12965;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Myosin Ie (Myosin Ic).
GN MYOIE OR MYOIC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]

SEQUENCE OF 1-245 FROM N.A.
RX MEDLINE=95018277; PubMed=7932763;
RA Bement W.M., Wirth J.A., Mooseker M.S.;
RT "Cloning and mRNA expression of human unconventional myosin-IC. A
RT homologue of amoeboid myosins-I with a single IQ motif and an SH3
RT domain.";
RL J. Mol. Biol. 243:356-363(1994).
CC -1- FUNCTION: Myosins are actin-based motor molecules with ATPase
CC activity. Unconventional myosins serve in intracellular movements.
CC Their highly divergent tails are presumed to bind to membranous
CC compartments, which would be moved relative to actin filaments (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.
CC -1- SIMILARITY: Contains 1 IQ domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

EMBL; U14391; AAA62667.1; -;
DR PIR; S53601;
DR HSP; P08799; IMND.
DR Genew; HGNC:7599; MYOIE.
DR MIM; 601479; -;
DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0003774; F:motor activity; TAS.
DR GO; GO:0008570; F:myosin ATPase activity; TAS.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00442; MYSC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50096; IQ; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Myosin Ie (Myosin Ic).
KW Multigene family.
FT DOMAIN 1 679 MYOSIN HEAD-LIKE.
FT DOMAIN 695 724 IQ.
FT DOMAIN 1052 1109 SH3.
FT DOMAIN 581 591 ACTIN-BINDING (POTENTIAL).
FT

FT NP BIND 112 119 ATP (POTENTIAL).
SQ SEQUENCE 1109 AA; 127040 MW; 0C1C74BAA3705905 CRC64;

Query Match 5.3%; Score 100; DB 1; Length 1109;
Best Local Similarity 22.5%; Pred. No. 10;
Matches 64; Conservative 41; Mismatches 91; Indels 88; Gaps 15;

QY 47 GADCAKFKSELEF-----KFNKALERPVTSKHSMGKTYGEHKKHLEFHDQYRELQRYA 102
DB 476 GADQTLQLQMQIGSHEFN-----SWNQGLIHHYAGKVSVDMDGFCERNR 523
QY 103 EEVGIPTASGMDMAVEFLHNLNVFFKVGSGDTNNFFLEKTKAKGRPMVSSGMQSM 162
DB 524 D---VLF---MD--LIELMQSSLEPFIK-----SLFPENLQADKKGRPTTAGSKIK-- 566
QY 163 DTMKQVQIVKPLNPNFCLOCTISAY-----PLQPEDVNLRVISEYQKLPPDIP 211
DB 567 ---KQANDIVSTL-----MKCTPHYIRCIKPNETKKPRDWESRVKHQVYGLKENIR 617
QY 212 IGVSGHETGIAISVAALGAKVLERHITLTK-----TWKSGDSHSASLEPGELAEVRSVR 267
DB 618 VRRAGY-----AYRIFQKFLQRYAILTKATWPSWQGEKQGVL-----HLQSVN 663
QY 268 LVERALGSPKQLLPCEMACNEKLGKSVVAKVPIEGTILTMDM 311
DB 664 M-----DSDQF-----QLGRSKVF-IPAKESLFLLEEM 690

RESULT 15

GPR BACME
ID GPR BACME STANDARD; PRT; 371 AA.
AC P2321;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Germination protease precursor (EC 3.4.24.78) (Spore protease) (GPR
DE endopeptidase) (Germination proteinase).
GN GPR.
OS Bacillus megaterium.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1404;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 16-31.
RC STRAIN=ATCC 12872 / QMRDC B1551;
RX MEDLINE=91100296; PubMed=1840582;
RA Sussman M.D., Setlow P.;
RA "Cloning, nucleotide sequence, and regulation of the Bacillus
RA subtilis gpr gene, which codes for the protease that initiates
RA degradation of small, acid-soluble proteins during spore
RA germination.";
RL J. Bacteriol. 173:291-300(1991).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE=98422459; PubMed=9748439;
RX Nessi C., Jedrejas M.J., Setlow P.;
RA "Structure and mechanism of action of the protease that degrades
RA small, acid-soluble spore proteins during germination of spores of
RA Bacillus species.";
RL J. Bacteriol. 180:5077-5084(1998).
RN [3]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF ZMOGEN P46.
RX MEDLINE=20323395; PubMed=10864493;
RX Ponnuraj K., Rowland S., Nessi C., Setlow P., Jedrejas M.J.;
RA "Crystal structure of a novel germination protease from spores of
RA Bacillus megaterium: structural arrangement and zymogen activation.";
RL J. Mol. Biol. 300:1-10(2000).
CC -1- FUNCTION: Initiates the rapid degradation of small, acid-soluble
CC proteins during spore germination.
CC -1- CATALYTIC ACTIVITY: Endopeptidase action with P4 Glu or Asp, P1
CC preferably Glu > Asp, P1', hydrophobic and P2' Ala.
CC -1- SUBUNIT: Homotetramer.
CC -1- DEVELOPMENTAL STAGE: GPR transcription occurs during sporulation
CC in forespore first by sigma-F and then by sigma-G.

```
CC -!- PTM: Autoproteolytically processed. The inactive tetrameric
CC zymogen termed p46 autoprocesses to a smaller form termed p41,
CC which is active only during spore germination.
CC -!- SIMILARITY: Belongs to peptidase family M63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55262; AAR22499.1; -.
DR PIR; A39198; A39198.
DR PDB; 1C8B; 03-MAY-01.
DR MEROPS; M63.001; -.
DR HAMAP; MF.00626; -.
DR InterPro; IPR005080; Peptidase M63.
DR Pfam; PF03418; Peptidase_U3; 1.
DR ProDom; PD041835; Peptidase_M63; 1.
DR TIGRFAMs; TIGR01441; GPR; 1.
KW Hydrolase; Protease; Zymogen; 3D-structure.
FT PROPEP 1 15
FT CHAIN 16 371 GERMINATION PROTEASE.
SQ SEQUENCE 371 AA; 40626 MW; 40ABE62ECA2F935D CRC64;

Query Match 5.2%; Score 98.5; DB 1; Length 371;
Best Local Similarity 20.9%; Pred. No. 3.4;
Matches 63; Conservative 44; Mismatches 120; Indels 75; Gaps 12;

QY 100 RYA--EEVGIFTASGDMEMAVE-----FLHELNV-----PFKVGSGDTNNFP--- 141
DB 71 RYVTLSEVGIREQDTEKQEEAMEEVFAKELNPFIKSLNIPDDASCLVGLGSLVTPDAL 130
QY 142 -----YLEKTAKGRPM--VISS--GMQSDMTMKQVYQIVKPLNPFC 180
DB 131 GPKAVDNLITRHLPELOPESVDGFRPVSAIVPGVMGNTGIETSDIIFGVVKVNPDI 190
QY 181 F-LQCTSAVPLQPDVNLAVISEYQKLPDIPFGYSGHETG---IAISVAVALGAKYLE 236
DB 191 IAI DALAARSI--ERVNATIQISDSGIHPGSGVGNKRKEISYETLGPVIAIGIPTWDA 248
QY 237 RHITLDKTWKS DSHASLEPGELAEVLRSVRLVERALGSPTKQLLPCEMACNEKLGKSVV 296
DB 249 VSI TSD-----TIDFILKHFGREMKEQKPSKSLP SGMTFGE----- 286
QY 297 AKVKIPEGTILTMDMLTVKVGPKAYPPEDIENLV-----GKKVLVTVEEDDTIMEEL 349
DB 287 -KKKLTEDDLFNEQRQTYLGMIGTLPDEEKRLIHEVLAPLGHNLMTPTKQVDMFIEDM 345
QY 350 VD 351
DB 346 AN 347
```

Search completed: September 13, 2004, 14:41:30
Job time : 28 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:38:59 ; Search time 32 Seconds
(without alignments)
579.179 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 1887

Sequence: 1 MPELELCGRWVGQHPFCF.....BEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	100.0	359	4	US-09-516-143A-4
2	336.5	17.8	346	4	US-09-495-406-21
3	113.5	6.0	343	4	US-09-107-532A-6711
4	100	5.3	547	4	US-09-648-004-10
5	93.5	5.0	389	4	US-09-134-000C-5846
6	93	4.9	375	4	US-09-252-991A-24216
7	93	4.9	868	2	US-08-907-166-2
8	93	4.9	868	4	US-09-391-340-2
9	92.5	4.9	753	4	US-09-543-681A-5022
10	92	4.9	310	4	US-09-498-520A-12
11	91.5	4.8	576	4	US-09-252-991A-25083
12	91	4.8	336	4	US-09-134-001C-3575
13	90	4.8	1211	4	US-09-134-001C-4820
14	89	4.7	831	4	US-09-269-861A-8
15	88.5	4.7	247	4	US-09-134-000C-5157
16	88.5	4.7	414	1	US-09-337-913-1
17	88.5	4.7	414	2	US-08-750-524-1
18	88	4.7	546	2	US-08-492-027A-8
19	87	4.6	679	3	US-08-913-942-15
20	87	4.6	679	4	US-09-268-347-26
21	87	4.6	1456	4	US-09-134-000C-6427
22	86.5	4.6	402	3	US-09-347-801-16
23	86.5	4.6	563	4	US-08-311-731A-142
24	86	4.6	519	4	US-09-489-039A-11591
25	85.5	4.5	382	1	US-07-768-286B-6
26	85.5	4.5	382	1	US-08-487-823B-3
27	85.5	4.5	382	2	US-08-997-040-3

28	85.5	4.5	382	2	US-09-203-237-3	Sequence 3, Appli
29	85.5	4.5	760	3	US-08-928-941D-29	Sequence 29, Appl
30	85.5	4.5	760	4	US-09-280-590A-29	Sequence 29, Appl
31	85.5	4.5	760	4	US-09-892-398-29	Sequence 29, Appl
32	85	4.5	789	4	US-09-134-000C-4939	Sequence 4939, Ap
33	85	4.5	1201	4	US-09-328-352-5890	Sequence 5890, Ap
34	84.5	4.5	303	3	US-08-928-941D-18	Sequence 18, Appl
35	84.5	4.5	303	4	US-09-280-590A-18	Sequence 18, Appl
36	84.5	4.5	303	4	US-09-892-398-18	Sequence 18, Appl
37	84.5	4.5	323	2	US-09-019-216-3	Sequence 3, Appli
38	84.5	4.5	323	4	US-09-249-241-3	Sequence 3, Appli
39	84.5	4.5	392	1	US-07-768-286B-4	Sequence 4, Appli
40	84.5	4.5	405	1	US-08-121-714-7	Sequence 7, Appli
41	84.5	4.5	405	1	US-08-477-108A-7	Sequence 7, Appli
42	84.5	4.5	405	2	US-08-477-112-7	Sequence 7, Appli
43	84.5	4.5	405	5	PCT-US93-08322-7	Sequence 7, Appli
44	84.5	4.5	415	1	US-07-911-531-19	Sequence 19, Appl
45	84.5	4.5	415	1	US-07-693-636A-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1

US-09-516-143A-4
; Sequence 4, Application US/09516143A
; Patent No. 6333182
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF0505PCT
; CURRENT APPLICATION NUMBER: US/09/516,143A
; CURRENT FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-516-143A-4

Query Match 100.0%; Score 1887; DB 4; Length 359;
Best Local Similarity 100.0%; Pred. No. 2.8e-193;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MPELELCGRWVGQHPFCFIIAETGQNHQGDLDVAKEMIRMAKCGADCAKFKSELEF	60
Db	1	MPELELCGRWVGQHPFCFIIAETGQNHQGDLDVAKEMIRMAKCGADCAKFKSELEF	60
Qy	61	KFNKALERPVTSKHSGKTYGEHKRHLEFSDQYRELQRYAEVGIFFTAGMDMAVE	120
Db	61	KFNKALERPVTSKHSGKTYGEHKRHLEFSDQYRELQRYAEVGIFFTAGMDMAVE	120
Qy	121	FLHELNVPPFFKVGSDTNNFPYLEKTAKKGRPMWISSQMSQMDTKQYQIVKPLNPNFC	180
Db	121	FLHELNVPPFFKVGSDTNNFPYLEKTAKKGRPMWISSQMSQMDTKQYQIVKPLNPNFC	180
Qy	181	FLOCTSAYPELOPEDNLRVISEYOKLPDIPIGYSGHETGTAISVAALCAKVLHERIT	240
Db	181	FLOCTSAYPELOPEDNLRVISEYOKLPDIPIGYSGHETGTAISVAALCAKVLHERIT	240
Qy	241	LDKTKGSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVAKVK	300
Db	241	LDKTKGSDHSASLEPGEALVRSVRLVERALGSPTKQLLPCMACNEKLGKSVAKVK	300
Qy	301	IPEGTILTMMDLTVKVGEPKAYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS	359
Db	301	IPEGTILTMMDLTVKVGEPKAYPPEDIFNLVGKKVLVTVEEDDTIMEELVDNHGKKIKS	359

RESULT 2

```
US-09-495-406-21
; Sequence 21, Application US/09495406
; Patent No. 6503744
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000110US
; CURRENT APPLICATION NUMBER: US/09/495,406
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 21
; LENGTH: 346
; TYPE: PR1
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of LOS biosynthesis locus)
US-09-495-406-21

Query Match          17.8%; Score 336.5; DB 4; Length 346;
Best Local Similarity 29.1%; Pred. No. 2.4e-27;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIQNHQGDLDVAKRMIRMAKCGADCAKFKSELEPKFNKALE-RPYTSHKSWGK 79
DB 19 VVEIIGINHMSLELAKIWDVAFTGCAKIIHQTHIVEDEMSKAARKVIPGNAKIS--- 75
QY 80 TYGEHRRHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFKVSGDNTN 139
DB 76 -IYEIMQKCALDYKDELALKEYTEKLGVLSTFFSRAGANRLMDGMVSFAFKIGSGECNN 134
QY 140 FVLEXTAKGRPMVLSSGQSDMTWKQVQIVKPLNPFCLQCTSAVPLQPEDVNLRV 199
DB 135 YPLIKHIAFKPMIVSTGNSTESIKPTVKIILDNEIIPVLMHTNLYTTPHNLRLNA 194
QY 200 ISEYQKLFDPIDIGYSGHETGIAISVAALGAKVLERHITLDTKWTGSDHSASLEPGE 259
DB 195 MLELKKEF-SCWGLSDHTDNLACLGAVALGACVLERHFTDSMHRSGPDIVCSDMTQAL 253
QY 260 AEL-VRSVRLVERALGSPKQLLPCMACNEKLGKSVAKVKIPGEGTILTMMLTVKVG 318
DB 254 KELIQSEQMAIMRGNESKAAKQEQVTTIDFASFVSVISIKDKKGEVLSDMNIWKRPG 313
QY 319 PKAYPPEDIFNLGKKVLVTVEEDDTIMEE 348
DB 314 LGGISAAEFENILGKKALRIENDTQLSYE 343

RESULT 3
US-09-107-532A-6711
; Sequence 6711, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC

US-09-107-532A-6711
; Sequence 6711, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Chen, Qiong
; APPLICANT: Thomas, Stuart
; APPLICANT: Nagarajan, Vasanth
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/252,553
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
```

```
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6711:
SEQUENCE CHARACTERISTICS:
LENGTH: 343 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...343
SEQUENCE DESCRIPTION: SEQ ID NO: 6711:
US-09-107-532A-6711

Query Match          6.0%; Score 113.5; DB 4; Length 343;
Best Local Similarity 22.4%; Pred. No. 0.0017;
Matches 65; Conservative 48; Mismatches 102; Indels 75; Gaps 14;

QY 10 GRWVGQHPFCIIAEIQNHQGDLDVAKRMIRMAKCGADCAKFKSELEPKFNKALER 69
DB 96 GSPVTWAGPCSI-----EG-LDQIRECARMAKAGGA-----KILRGGAFAK 134
QY 70 PYTSKHSWKTGYEHRHLEFSDHQRELQRYAEVGIFFTAGMDMAVEFLHNLNVPF 129
DB 135 PRTPSYAF---QGLEEEGLKY-----IRQADEFQMKVITEVMDGHDIMVAEYS-DI 183
QY 130 FKVGSDTNFPVLEKTAKKGRPMVLSSGQSDMTWKQ-----VYQIVKPLNPFCLQ- 183
DB 184 LQIGARNMQNFKLSAVGKTPGKVGKRGISG--TINELNAAEYIAVEDKSP-VIPIER 240
QY 184 -----CTSAVPLQPEDVNLRVISEYQKLFDPIDIPIGYSGHETGI-----AISV 225
DB 241 GIRTYETATNTFDLSAVPLMKKLTHTFPVIVD-----PSHGTGIEWELVPPMAR 288
QY 226 AVALGA--KVLRRHITLDTKWTGSDHSASLEPGEALAEVRSVRLVERAL 273
DB 289 AGVASGADGMIVBIHPDPANAW--SDGPQSLNEXKTYSRMMKVEAIMKEAM 336

RESULT 4
US-09-648-004-10
; Sequence 10, Application US/09648004
; Patent No. 6498242
; GENERAL INFORMATION:
; APPLICANT: Chen, Qiong
; APPLICANT: Thomas, Stuart
; APPLICANT: Nagarajan, Vasanth
; TITLE OF INVENTION: BIOLOGICAL METHOD FOR THE PRODUCTION OF ADIPIC ACID AND
; INTERMEDIATES
; FILE REFERENCE: CL-1341-A
; CURRENT APPLICATION NUMBER: US/09/648,004
; CURRENT FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/252,553
; PRIOR FILING DATE: 1999-02-19
; NUMBER OF SEQ ID NOS: 32
```



```

; CURRENT APPLICATION NUMBER: US/08/907,166
;
; CURRENT FILING DATE: 1997-08-06
;
; NUMBER OF SEQ ID NOS: 12
;
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 2
;
; LENGTH: 868
;
; TYPE: PRT
;
; ORGANISM: Ammonifex degensii
;
; US-08-907-166-2

```

Query Match	4.9%;	Score 93;	DB 2;	Length 868;
Best Local Similarity	19.8%;	Pred. No. 1.2;		
Matches	72;	Conservative 55;	Mismatches 102;	Indels 134; Gaps 17;

Qy	51	AKFOKSELEPKFNKALERP	YTSKSHGKGTVEGKHRL	FLFESHDOYRELQRYAE	EVGIFFT	110		
Db								
Qy	57	AAFDKSKTF	-----RHALAET	Y--KAHRPATPDLRQ	QFNLIKEV	95		
Db								
Qy	111	ASGDEMAVEFLH	ELNVPFFKVGSGDTNNF	--PYLEKTA	KGRPMWISSGMSQSDMTKQV	168		
Db								
Qy	96	-----LTALN	VPVVEKEGFEADDL	IGTLVDRAEKEGMQCL	IVTG--DLDAL---	139		
Db								
Qy	169	YQIVKPLNPNCF	LOCTSA	YPLQEP	-----DYNLRVISEYQK	214		
Db								
Qy	140	-QLVSP	LTVTVMKRGISE	IAVFNAEVR	RGFTVPQLPOLPQ	KALAGDASNIPLP	195	
Db								
Qy	215	SGHETGIAIS	VAALGAKVL	ERHITLDTKWGSDH	SASLFPGLAE	LVRSVRIVLRALG	274	
Db								
Qy	196	-----GIGPKTA	-----SRLLQSHQSLEK	LE---	SKEFPFAKL	RETLE--RHKEBAVL	239	
Db								
Qy	275	SPTQLQLPC	EMACNEKL	-----GKSVAKVKI	-----PRGTLTWDML	312		
Db								
Qy	240	AKKLALTRD	VPLEEEITRP	WPGPNILATLV	SFRLEFRTLAKR	FLFEPPEARLLSASGL	299	
Db								
Qy	313	T-----VKVGRPK	-----AYPP	-----EDIFN	IVGK	333		
Db								
Qy	300	TPSAVRVKV	RPBELLER	GEELGRQEP	AAALAYPVL	RRKATSSFLA	LCGKEKVFLEGP	359
Db								
Qy	334	KVL	336					
Db								
Qy	360	EVL	362					
Db								

```

RESULT 8
US-09-391-340-2
; Sequence 2, Application US/09391340A
; Patent No. 6492511
; GENERAL INFORMATION:
; APPLICANT: Callien, walter
; APPLICANT: Mather, Eric
; TITLE OF INVENTION: ISOLATION AND IDENTIFICATION OF NOVEL POLYMERASES
; FILE REFERENCE: 09010/027001
; CURRENT APPLICATION NUMBER: US/09/391,340A
; CURRENT FILING DATE: 1999-09-07
; EARLIER APPLICATION NUMBER: US 08/907,166
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 868
; TYPE: PRT
; ORGANISM: Ammonifex degensii
US-09-391-340-2

```

```
Query Match          4.9%; Score 93; DB 4; Length 868;  
Best Local Similarity 19.8%; Pred.No.1.2;  
Matches             72; Conservative 55; Mismatches 102; Indels 134; Gaps 17  
  
QY      51 AKFQKSELLEFPNKRKALERPYYTKSHGWTGYGEHRHLEFSHDQVRELQRVAEVEWGIIFFT 110  
         ||| :||  
DB       57 AAFDKSKTTF-----RHAALEY---KAHRPATPDELRQQDFNLKIKEY---- 95  
  
Ov     111 ASGMDEMAYEFELHNVPFFKFVGSGDGNNF--PYLEKTAKRGRPMVISGGMSMDTMKQV 168
```

[illegible]

RESULT 9

```

US-09-543-681A-5022
; Sequence 5022, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEIN
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5022
; LENGTH: 753
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5022

```

Query Match	4.9%;	Score 92.5;	DB 4;	Length 753;	
Best Local Similarity	20.8%;	Pred. No. 1.1;			
Matches	69;	Conservative 58;	Mismatches 105;	Indels 99;	Gaps 18
QY	117	MAVEFLHLENVPFFKVGSGDITNNFPY-----LEKTAKKGRPVVISGMSQ	161		
		: : : : : :	: : : : : :		
Db	152	IAKDFLHDLNFFVIYDQKEYTLPIYGEQSWSEKKRTLLELEKLABEGNKEAVYOYLEN	211		
		: : : : : :	: : : : : :		
QY	162	MDTWKQ-----VYQIVKPLNPNCFFLOCTSAYPLOQPDVNLRVISEY-----	203		
		: : : : : :	: : : : : :		
Db	212	KDSVKQSKTKNMIRPPIFEKDPKMYLK---YKLSQDINLLFDSAYLNVEATYLLY	267		
		: : : : : :	: : : : : :		
QY	204	-----QKLFP-----DPIGYS-----GHETGAISVAV-----ALGA-----KVLERH	238		
		: : : : : :	: : : : : :		
Db	268	QIYKGDKNYPLQENSFLANIYLLKKSADLAHHDGLIKIIEELNNSNTLSSNYFRNLLEKY	327		
		: : : : : :	: : : : : :		
QY	239	I-TLDTKWGSDH-----SASLEP-GELAELVRSVRLVERALG-----SPTKQLLPCEMAC	287		
		: : : : : :	: : : : : :		
Db	328	IDTLALKYPNSQOAMLALANVYLKPNSSFYNFEEKALKLVKEAYNTQPSPESKULLAKLYS	387		
		: : : : : :	: : : : : :		
QY	288	N-EKLGKSVAKVIPEGTILTMDLT-----VKV-----GPKRAYPEDIFNLVGKKV	335		
		: : : : : :	: : : : : :		
Db	398	NSEGVHQRKAVSFLKENI-TNDKLTGKSORELVKIYDFGASDYLKKEEIVNLR	446		
		: : : : : :	: : : : : :		
QY	336	LV-----TVEED-----DTIMEELVDNH	353		
		: : : : : :	: : : : : :		
Db	447	INKKSAGTFNQNYSIAHFYADULLLEEDVANN	477		


```

QY      9  GRW---VGGHPFII--AEIQNHQGDLD-----VAKRMTIRMAKECCADCAKFO  54
      |||
Db     138  PTGHPAFVAGAHGAALSIAPGSFGPTWGDLSERRGWPAVARRMASFPCLRILAILAREQ  197
      :|:|:|
QY     55  KSELEFENRKALERPYSKHSWGKTYGEGHKRHLFSDHOYRELQRYAEVEVGFFFTASGM  114
      :|:|:|
Db     198  DAVHE-----ERPGPERRA-----SCHLYLSF-----VLLPNGF  228
      :|:|:|
QY     115  DEMAVEFELHNLVPPFKVSGDGTNNPYLEKTAKGRPMVISGMSQMDTMKVQYQIVK-  173
      :|:|:|
Db     229  R-----MKAAGGRSSRRPGVLSVPI-RPVLPARGMYAMSVTDASFFLIINW  273
      :|:|:|
QY     174  ---PLNPNFCLOCTSA---PLQPEDVNLRVISEYOKLPDPDIPGVS  215
      :|:|:|
Db     274  SKAMPSLRQLMYFACVABEYSGISQAEEVLHVVSQPSLSRQIALEVHIGBALFVRGPAGVS  333
      :|:|:|
QY     216  GHETGIAISVAVAL--GAKVLERHI-----TLDKTWKG-----  247
      :|:|:|
Db     334  LTEGVRVHVCRAVKIIAEAKALTRDIGDDRRQLHIGFSRYLDPGWSGELARQLKEARVC  393
      :|:|:|
QY     248  SDHSASLEPG-ELAEVRSVKLVRERALGSPTKQLLPCEM-----ACNEKLGKSVV  296
      :|:|:|
Db     394  SAVGFSWAPSLDLLSLVRESSLDAAIIGAPPKVVRVEYELMEIGSYEHVIALSETLYRRLA  453
      :|:|:|
QY     297  AKVKIPGTLTMDMLTVKVGCEPKAYPPEDIF  328
      :|:|:|
Db     454  -----PWSLNDL-----IGEPLLVPKDSL  474
      :|:|:|

RESULT 12
US-09-134-001C-3575
; Sequence 3575, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3575
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3575

Query Match      4.8%; Score 91; DB 4; Length 336;
Best Local Similarity 21.7%; Pred. No. 0.43;
Matches 52; Conservative 32; Mismatches 64; Indels 92; Gaps 10

QY      141  PYLEKTAKGRPMVISGMSQMDTMKVQYQIVKPLNPNFCLOCTSAYPLOPEDVNL---  197
      :|:|:|
Db     27  PHLE---NKKVLIGA-----TTSNPYHAINPAIRSRA--QIFELYPLDQDDIRLALD  74
      :|:|:|
QY     198  ---RVISEYOKLPDPDIPGYSGHETGIAISVAVALGAKVL-----ERHITLD  242
      :|:|:|
Db     75  RAINDKERGLSTYHPIVDEDAIEYFTSQSDVRSALNALEVLASHIHGEENERHITLD  134
      :|:|:|
QY     243  ---KTKGSDHSASL-----EPGELABELVRSVR  267
      :|:|:|
Db     135  DAKDCLQKGAFAVSDKGDMDHYDVMSAFQKSIRGSDVNAALHYLARLIEAGDLPPIVRRL  194
      :|:|:|
QY     268  LV---EHALGSPTKQLLPCEMACNEKLGKSVAKVIPEGTILTMQMLTVKVGEPKAYPP  324
      :|:|:|
Db     195  VISYEDVDGLASP-----NAGQRTLLAIQSAE-----RLGFPPEARIP  230
      :|:|:|

```

```
RESULT 13
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4820
; LENGTH: 1211
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820

Query Match 4.8%; Score 90; DB 4; Length 1211;
Best Local Similarity 19.2%; Pred. No. 4.3;
Matches 75; Conservative 69; Mismatches 144; Indels 102; Gaps 21;

QY 26 GQHQQDLDVAKRMIRMAKECCACAKFOKSELEFKFNKALERPPTSXKSWGKTYGEHK 85
Db 297 GKRQNDYDIEKLNVELVK-----ATENYEOGSKLN--VLEERKKNQSENNARYEEL 348
QY 86 RHLEFSDH-----OYRELQRYAEVGIFFFTASGMDMAVEFFLHENV 127
Db 349 DNLQSDIQSIKNEKAQEKLLADLNKQKQLNKEVQELSELISLYIS--DQHDKLEBEIKN 406
QY 128 PFFKVGSGDT---NNPPYLEKT-----AKGRPMVTSQMGQMDTMQVYQIVKPLNPNP 179
Db 407 SYTLMSQGVNNDIRFLEHTINENEAKKSR--LDSRLVEAPNQLKDIQNNITQKEY 464
QY 180 CFLOCTSAVPLQPEDVNL-----RVISEYQKLPDIPIGYSGHETGTAISVAAV 228
Db 465 ----QSSKSKMEKVEQNIQQLQQLTDSKELLSEYENKL-----YQARYNEKLKSRID 514
QY 229 ALGAKVLE-----RH1--TLDKTKGSDHSASLE-----PGLAELVRSVRLVBRALG 274
Db 515 SLATQEDTYTFENGKHLKAKDKELRGI-HGAVAEVINVPSEMTQ-----AIEALG 567
QY 275 SPTKOLLPCMACNEKLGKSVVAKVK---IPEGTILTMDLTVKVGEPKAYPPEDIPNLV 331
Db 568 ASLQHV1---VDNEKDGROAIQYLKQKQGLGRATFLPLNVI-----QPR-HVAADIKDVA 617
QY 332 -GKKVLTVBED-----DTIMEELVDN 352
Db 618 RGSQGFINTASDAINVSAXYQNIENLLGN 647

RESULT 14
US-09-269-861A-8
; Sequence 8, Application US/09269861A
; Patent No. 6468775
; GENERAL INFORMATION:
; APPLICANT: Ankenbauer, Waltraud
; APPLICANT: Markau, Ursula
; APPLICANT: Svetlichny, Vitaly
; APPLICANT: Schmitz-Aghueglian, Gudrun
; APPLICANT: Reiser, Astrid
; APPLICANT: Angerer, Bernhard
; APPLICANT: Ebenbichler, Christine
; APPLICANT: Laue, Frank
; APPLICANT: Bonch-Osmolovskaya, Elizaveta
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASE FROM CARBOXYDOTHERMUS HYDROGENOFORMAN
; FILE REFERENCE: 4494
; CURRENT APPLICATION NUMBER: US/09/269,861A
; CURRENT FILING DATE: 1999-11-22
```

```
; PRIOR APPLICATION NUMBER: PCT/EP97/05391
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: EP/96115873.0
; PRIOR FILING DATE: 1996-10-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Carboxydotherrmus hydrogenoformans
US-09-269-861A-8

Query Match 4.7%; Score 89; DB 4; Length 831;
Best Local Similarity 22.3%; Pred. No. 3;
Matches 72; Conservative 38; Mismatches 105; Indels 108; Gaps 15;

QY 54 QKSELEFKFNKALERPPTSXKSWGKTYGEHKHLEFSDHQYRELQRYAEVGIFFT--- 110
Db 216 RXDELKPKLREKLAHEHLAKISKQIATILRIPLEISLEDLKVKPEPVEEYVAKLFLHLE 275
QY 111 -----ASGMDMAVEFFLH---ELNVPF---FKVSGDGTNNFPYLEKTAK 148
Db 276 FKSFLKEIEPKIKKEYQEGKDLVQVETVETEGQIAVVSDFVDDGEKTKYSLDLRNE 335
QY 149 -----KGRPMVTSQMGQMDTMQVYQIV--KPLN--PNFCFLOCTSAVPLQ- 192
Db 336 IBEIFRNKKII-----TDDAKGIYHVCLEKGLTFPEVCFDARIAAYVLNPADQNPGLK 388
QY 193 -----PDVNLRV-----ISEYQKLPDI-----PIGVSGHETGI 221
Db 389 GYLYKYDLPVYEDVSLNIRGLFYLKEMMRKIFEQQERLFYEIEIPLFVLVLAQMEHTGI 448
QY 222 AISVAVALGAKVLERHITLDTKTKGSDHSASLEGE-LAELVRSVRLV---ERALGSPT 277
Db 449 QVDREAL-----KEMSELGEQIEELIREIYVLAGEEFLNLS- 486
QY 278 KOLLPCMACNEKLGKSVVAKVK 300
Db 487 RQL---GVILFEKGLGIPVIKTKK 506

RESULT 15
US-09-134-000C-5157
; Sequence 5157, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5157
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5157

Query Match 4.7%; Score 88.5; DB 4; Length 247;
Best Local Similarity 22.0%; Pred. No. 0.48;
Matches 49; Conservative 35; Mismatches 84; Indels 55; Gaps 11;

QY 34 DVAKR-----MIRMAKECCACAKFOKSELEFKFNKALERP-----YTSKHSWG 78
Db 44 EVAKHCFSSPSSIRLAKNLNS---GYNELIYKLEKAHFSQPIPETAPSPETTFNEFC 99
QY 79 KTYGEHKRHL--EFSHDQYRELQRYAEVGIFFFTASGMDMAVEFFLHENVFPKVGSGD 136
Db 100 RLLAKKSHLFLVILGHDFSRHLAAYISEV---FNFHGIPTITATTYTHSIN-----SQN 149
```

QY 137 TNNE-----PYLEKTA-----KGRBPMVISSQMOMDMKQVYQIVKPLNENPCF 181
Db 150 NQNFLEFIILSHSGEEKYLLAKETALLAKEKKHPIISFVGAKN-STLGRDLADLVFSTDYSYSPF 208
QY 182 LOCTSAYPLQPEDVNLRVISEYQKLFPPDIPIGYSGHETGIAIS 224
Db 209 ----STSVAPQPMFFGQTLITFEALI-----CAYLNHEDSIPIS 243

Search completed: September 13, 2004, 14:45:00
Job time : 34 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:43:50 ; Search time 130 Seconds
(without alignments)
885.597 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MPLELELCGRWVGQHPCF.....BEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1335176 seqs, 320689617 residues

Total number of hits satisfying chosen parameters: 1335176

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	1887	100.0	359	9	US-09-984-205-4
2	1887	100.0	359	9	US-09-930-440B-6
3	1887	100.0	359	16	US-10-759-277-4
4	1883	99.8	359	15	US-10-264-237-2597
5	544.5	28.9	338	9	US-09-767-041-24
6	524	27.8	311	14	US-10-156-761-10910
7	510.5	27.1	346	9	US-09-930-440B-8
8	381	20.2	123	14	US-10-106-698-6926
9	336.5	17.8	346	9	US-09-816-028A-35
10	336.5	17.8	346	14	US-10-303-161-35
11	336.5	17.8	346	14	US-10-303-118-35
12	336.5	17.8	346	14	US-10-303-128-35
13	336.5	17.8	346	14	US-10-303-134-35
14	336.5	17.8	346	14	US-10-303-162-35
15	303	16.1	340	10	US-09-882-227-490

16	287	15.2	340	12	US-10-335-977-7626	Sequence 7626, Ap
17	283.5	15.0	314	12	US-10-335-977-7625	Sequence 7625, Ap
18	122.5	6.5	205	12	US-10-321-396C-4	Sequence 4, Appli
19	119	6.3	352	15	US-10-369-493-18790	Sequence 18790, A
20	117.5	6.2	275	12	US-10-321-396C-2	Sequence 2, Appli
21	116	6.1	352	15	US-10-369-493-20092	Sequence 20092, A
22	113.5	6.0	333	12	US-10-282-122A-57496	Sequence 57496, A
23	113	6.0	285	15	US-10-369-493-18119	Sequence 18119, A
24	110	5.8	135	12	US-10-321-396C-6	Sequence 6, Appli
25	110	5.8	265	15	US-10-369-493-21568	Sequence 21568, A
26	109	5.8	337	15	US-10-369-493-9746	Sequence 9746, Ap
27	108.5	5.7	335	15	US-10-369-493-20181	Sequence 20181, A
28	108.5	5.7	351	15	US-10-369-493-2631	Sequence 2631, Ap
29	108.5	5.7	633	12	US-10-282-122A-63506	Sequence 63506, A
30	108	5.7	338	15	US-10-369-493-2930	Sequence 2930, Ap
31	107.5	5.7	337	12	US-10-282-122A-51516	Sequence 51516, A
32	107	5.7	276	15	US-10-369-493-22853	Sequence 22853, A
33	107	5.7	330	12	US-10-282-122A-67597	Sequence 67597, A
34	103.5	5.5	329	15	US-10-369-493-13704	Sequence 13704, A
35	101.5	5.4	335	15	US-10-369-493-8945	Sequence 8945, Ap
36	101.5	5.4	337	12	US-10-282-122A-53304	Sequence 53304, A
37	101.5	5.4	341	12	US-10-282-122A-57228	Sequence 57228, A
38	101	5.4	1175	15	US-10-369-493-6504	Sequence 6504, Ap
39	100	5.3	547	14	US-10-272-419-10	Sequence 10, Appl
40	100	5.3	547	14	US-10-230-026-20	Sequence 20, Appl
41	100	5.3	1104	16	US-10-322-281-778	Sequence 778, Ap
42	100	5.3	1108	12	US-10-087-192-1206	Sequence 1206, Ap
43	100	5.3	1189	12	US-10-282-122A-70920	Sequence 70920, A
44	99.5	5.3	702	16	US-10-437-963-169700	Sequence 169700,
45	98	5.2	715	16	US-10-437-963-120663	Sequence 120663,

ALIGNMENTS

RESULT 1

US-09-984-205-4
; Sequence 4, Application US/09984205
; Patent No. US20020137175A1
; GENERAL INFORMATION:
; APPLICANT: Coleman, Timothy A. et al.
; TITLE OF INVENTION: Human Glycosylation Enzymes
; FILE REFERENCE: PF50SD1
; CURRENT APPLICATION NUMBER: US/09/984,205
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: PCT/US00/05325
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/516,143
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/122,409
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-984-205-4

Query Match 100.0%; Score 1887; DB 9; Length 359;
Best Local Similarity 100.0%; Pred. No. 1.7e-182;
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPLELELCGRWVGQHPCFITAEIGNQHGDLVDVAKRMIRMAKCGADCAKTKSELEF	60
Db	1	MPLELELCGRWVGQHPCFITAEIGNQHGDLVDVAKRMIRMAKCGADCAKTKSELEF	60
QY	61	KFNKALERPYSKHSWGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE	120
Db	61	KFNKALERPYSKHSWGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE	120
QY	121	FLHMLNVPFKVSGSDTNFPVLEKTKAGRPVIVSSGMSMDTMKVQIVKPLPNFC	180

Db 121 FLHNLNVPFFKVGSGDTNNFYLEKTAKKGRPMVSSGMSQMDTMKQYQIVKPLNPNFC 180
 QY 181 FLQCTSAYPELVNLRVISEYQKLPDPIPIGYSGHETGTAISVAVALGAKVLERHIT 240
 Db 181 FLQCTSAYPELVNLRVISEYQKLPDPIPIGYSGHETGTAISVAVALGAKVLERHIT 240
 QY 241 LDKTWKGDHSDHSLPGLAEALVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK 300
 Db 241 LDKTWKGDHSDHSLPGLAEALVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 2
 US-09-930-440B-6
 ; Sequence 6, Application US/09930440B
 ; Patent No. US20020142386A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Betenbaugh et al.
 ; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
 ; FILE REFERENCE: PF509P2
 ; CURRENT APPLICATION NUMBER: US/09/930,440B
 ; CURRENT FILING DATE: 2001-08-16
 ; PRIOR APPLICATION NUMBER: 60/227,579
 ; PRIOR FILING DATE: 2000-08-25
 ; PRIOR APPLICATION NUMBER: 09/516,793
 ; PRIOR FILING DATE: 2000-03-01
 ; PRIOR APPLICATION NUMBER: 60/169,624
 ; PRIOR FILING DATE: 1999-12-08
 ; PRIOR APPLICATION NUMBER: 60/122,582
 ; PRIOR FILING DATE: 1999-03-02
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; TYPE: PRT
 ; LENGTH: 359
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-09-930-440B-6

Query Match 100.0%; Score 1887; DB 16; Length 359;
 Best Local Similarity 100.0%; Pred. No. 1.7e-182;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELELCPRWVGQHPCHFIIEIGQNHQGLDVAKRMRMAKEGADCAKFKQKSELEF 60
 Db 1 MPELELCPRWVGQHPCHFIIEIGQNHQGLDVAKRMRMAKEGADCAKFKQKSELEF 60
 QY 61 KFNKALERPYSKHSWGKTYGEHKLHLEFSDHQYRELQRYAEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPYSKHSWGKTYGEHKLHLEFSDHQYRELQRYAEVGIFFTAGMDMAVE 120
 QY 121 FLHNLNVPFFKVGSGDTNNFYLEKTAKKGRPMVSSGMSQMDTMKQYQIVKPLNPNFC 180
 Db 121 FLHNLNVPFFKVGSGDTNNFYLEKTAKKGRPMVSSGMSQMDTMKQYQIVKPLNPNFC 180
 QY 181 FLQCTSAYPELVNLRVISEYQKLPDPIPIGYSGHETGTAISVAVALGAKVLERHIT 240
 Db 181 FLQCTSAYPELVNLRVISEYQKLPDPIPIGYSGHETGTAISVAVALGAKVLERHIT 240
 QY 241 LDKTWKGDHSDHSLPGLAEALVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK 300
 Db 241 LDKTWKGDHSDHSLPGLAEALVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 4
 US-10-264-237-2597
 ; Sequence 2597, Application US/10264237
 ; Publication No. US20040009491A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Birse et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PA131P1
 ; CURRENT APPLICATION NUMBER: US/10/264,237
 ; CURRENT FILING DATE: 2002-10-04
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450
 ; PRIOR FILING DATE: 2001-05-18
 ; PRIOR APPLICATION NUMBER: US 60/205,515
 ; PRIOR FILING DATE: 2000-05-19
 ; NUMBER OF SEQ ID NOS: 2876
 ; SOFTWARE: PatentIn Ver. 3.1
 ; SEQ ID NO 2597
 ; LENGTH: 359
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; ORGANISM: Homo sapiens
 US-10-264-237-2597

RESULT 3
 US-10-759-277-4
 ; Sequence 4, Application US/10759277

```
Query Match          99.8%; Score 1883; DB 15; Length 359;
Best Local Similarity 99.7%; Pred. No. 4.4e-182;
Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPLELELCGRVWGQHQHCFIIAIEIGQNHQDLDVAKRMIRMAKCEGADCAKFKSELEF 60
DB 1 MPLELELCGRVWGQHQHCFIIAIEIGQNHQDLDVAKRMIRMAKCEGADCAKFKSELEF 60

QY 61 KPNRKALERPYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
DB 61 KPNRKALERPYSKHSWGKTYGHEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120

QY 121 FLHNLVFFKVGSDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPLNPNFC 180
DB 121 FLHNLVFFKVGSDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPLNPNFC 180

QY 181 FLOCTSAVPLQPEDVNLRAVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHIT 240
DB 181 FLOCTSAVPLQPEDVNLRAVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHIT 240

QY 241 LDKTWKSDHSALEPGLABELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
DB 241 LDKTWKSDHSALEPGLABELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300

QY 301 IPEGTILTMMDLTVKVGEPKAYPPEDIENLYGKVLVTVVEEDDTIMEELVDN 359
DB 301 IPEGTILTMMDLTVKVGEPKAYPPEDIENLYGKVLVTVVEEDDTIMEELVDN 359

RESULT 5
US-09-767-041-24
; Sequence 24, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 24
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: CPS2p
US-09-767-041-24

Query Match          28.9%; Score 544.5; DB 9; Length 338;
Best Local Similarity 36.0%; Pred. No. 3.2e-46;
Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;

QY 20 FIIAIEIGNHQDLDVAKRMIRMAKCEGADCAKFKSELEFKNRKALERPYSKHSWGK 79
DB 3 YIIAIEICNHGNDVHLARKMVEVAVDCGVDAKVFQTEKADLLISKYAPKAEY-QKITGE 61

QY 80 TYG--EHRKHLFSDQYRELQRYAEVGIFFTAGMDMAVEFLHNLVFFKVGSDT 137
DB 62 SDSQLEWTRRLLELSFEYLDLRDYLEKGVDVFTPEDESLDFLSTDMPVYKIPSGEI 121

QY 138 NNFPYLEKTKKGRPMVLSGMSQMDTMKVYQIVKPLNPN----FCFLOCTSAVPLQPE 193
DB 122 TNLPLYLEKIGRQAKKVLSTGMVMD--EIHQAVKILQENGTTDISILHCTTETPTYP 178
```

```
QY 194 DVNLRAVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHITLDKTWKGSDHSA 253
DB 179 ALNVLNLTKEFPNLTIGYSDHSGVEVPITAAAMGAELTEKFTLIDNEMEGPDHKAS 238

QY 254 LEPLGELABELVRSVRLVERALGSPTKQLLPCEMACNEKL--GKSUVAKVKIPGTTILTM 311
DB 239 ATPDILALAVKGVRLVEQSLGKFEKE--PEEVEVRNKIVAEKSIYAKKALAKGEVFTEN 296

QY 312 LTVKVGEP-KAYPPEDIENLYGKVLVTVVEEDDTIMEELVDN 352
DB 297 ITVK--RPGNGISPMWYKVLGVQSEQDPEEDQNICHSAFEN 336

RESULT 6
US-10-156-761-10910
; Sequence 10910, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10910
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10910

Query Match          27.8%; Score 524; DB 14; Length 311;
Best Local Similarity 38.9%; Pred. No. 3.4e-44;
Matches 122; Conservative 53; Mismatches 115; Indels 24; Gaps 7;

QY 14 GGOHPCFIIAIEIGQNHQDLDVAKRMIRMAKCEGADCAKFKSELEFKNRKA--LERPY 71
DB 15 GPGHPVYVVGIEGINHNGELENAFKLIDAAEAGCDVAKFKQKRTPEICTPRDQWDIER-- 72

QY 72 TSKHSWGK-TYGEKRLHLEFSDQYRELQRYAEVGIFFTAGMDMAVEFLHNLVDPFF 130
DB 73 --DTPWGRMTYIDYRHRVFEGEDEYRQIDEYAKSKNIDWFASPDWTEAVAFLEKFDIPAH 130

QY 131 KVGSGDTNNPYLEKTAKKGRPMVLSGMSQMDTMKVYQIVKPL-NPNFCFLOCTSAYP 189
DB 131 KVASASITDELLEALRGTRTVILSTG---MSTPKQIRHAEVGLGSDNILLCHATSYP 187

QY 190 LQPEDVNLRAVISEYQKLFDPDIPIGYSGHETGTAISVAVALGAKVLERHITLDKTWGS 249
DB 188 AKABELNLRVINTLQAEYVNPVIGYSGHETGLQTTLAVALGATFVERHITLDRAMWGS 247

QY 250 HSASLEPGLAEELVRSVRLVERALGSPTKQLLPCEMACNEKLGK--SVVAKVKIPEGTIL 307
DB 248 QAAASVEPQGLTRIVRDIRTIEASLGDGVKKVYESLGPMKKLRRVPGVAAEAI----- 301

QY 308 TMDMLTVKVGEPKA 321
DB 302 -----AAAGPEVA 310

RESULT 7
US-09-930-4408-8
; Sequence 8, Application US/09930440B
; Patent No. US20020142386A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Betenbaugh et al.
; TITLE OF INVENTION: Engineering Intracellular Sialylation Pathways
; FILE REFERENCE: PF509P2
; CURRENT APPLICATION NUMBER: US/09/930,440B
; CURRENT FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/227,579
; PRIOR FILING DATE: 2000-08-25
; PRIOR APPLICATION NUMBER: 09/516,793
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/169,624
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: 60/122,582
; PRIOR FILING DATE: 1999-03-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-930-440B-8

Query Match      27.1%; Score 510.5; DB 9; Length 346;
Best Local Similarity 36.3%; Pred. No. 9.4e-43;
Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FTIAEIQNHQDLDVAKRMIRMAKECGADCAKFKQSELEFKENRKALERPYSKHSWG- 78
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 YIVAEIGCNHGSVDIAREMILKAKEAGYNAVKFQTFKADKLISAIAPKAERYQIKNTGEL 64
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 79 KTYGEKRLHLEFSDQYRELQRYAEVGIFFFTASGMDENAVEFLHNLNVPFFKVGSGDTN 138
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 ESQLEMTKLEMKYDYLHLMYAVSLNDVSTPDESDIDPLASLKOKIKWIPSGELL 124
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 139 NFPYLEKTAKGRP---MVISGMSQMDTMKQYQIV---KPLNPNFCLQCTSAYPLOP 192
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 NLPYLEKIAKLPIPKKIIISTGMATIDEIKQSVSIFINNKPVGNIITILHCNTEYPTPF 184
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 193 EDVNLRVISEYQKLPDPIGYSCHTGTATSAVAALGAKVLERHITLDTKWKSGDHA 252
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 EDVNLNAINDLKKHFFKNIGFSDHSGFYAAAPVPGITFIEKHTLTKDKNSGDPHLA 244
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 253 SLEPGELAEIIVRSVRVLRALGSPTKQLLPCEMACNEKLGKSVVAKVKIPEGTILTMML 312
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 STEPDELKHLCTGRCVCEKSLGNSKVVYASERKNKIVARKSIIAKTEIKKGEVTFSEKNI 304
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVEEDDTIMEELV 350
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 305 TTK--RPGNGISPMWYNLLGK-----IAEQDFIPDELI 336
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-10-106-698-6926
; Sequence 6926, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6926
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (83)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (87)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (111)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (112)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (121)-
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6926
```

Query Match 20.2%; Score 381; DB 14; Length 123;
Best Local Similarity 93.7%; Pred. No. 2.6e-30;
Matches 74; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 183 QCTSAYPLOPQEDVNLRVISEYQKLPDPIGYSCHTGTATSAVAALGAKVLERHITLD 242
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 24 KCTSAYPLOPQEDVNLRVISEYQKLPDPIGYSCHTGTATSAVAALGAKVLERHITLX 83
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 243 KTWKSGDHSASLEPGELAE 261
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTWKSGDHSASLEPGELGE 102
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9

US-09-816-028A-35
; Sequence 35, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-09-816-028A-35

Query Match 17.8%; Score 336.5; DB 9; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHOGDLDVAKRMIRMAKECGADCAKFKQSELEFKENRKALERPYSKHSWGK 79
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 VPEIGININGSLELAKIWDAAFPSTGAKIHKQTHIVEDEMSKAACKVTPGNAKIS--- 75
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 80 TYGEKRLHLEFSDQYRELQRYAEVGIFFFTASGMDENAVEFLHNLNVPFFKVGSGDTNN 139
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 76 -IVEINQKCALDYKDELALKEYTEKLGVLVLPFSGRAGANRLDEMGVSFAFKIGSGCEN 134
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 140 FVPLETKAKGRPMVLISSGMSQMDTMKQYQIVKPLNPNFCLQCTSAYPLOPQEDVNLRV 199
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Wed Sep 15 10:55:31 2004

us-10-759-277-4.rapb

Page 5

```

Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEP-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVPIEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIFNLGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 10
US-10-303-161-35
; Sequence 35, Application US/10303161
; Publication No. US20030148459A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,161
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-161-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKRMIRMAKCGADCAKFKQSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHNGSLELAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIPGNAKIS--- 75
Qy 80 TYGEKRLHLEFSDHQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEIMQKCALDYKDELALKEYTEKGLVLTSTPFSRAGANRLDMGVSAFKIGSGECNN 134
Qy 140 FPYLEKTAKGRPMVSSGMQSDMTKQVQIVKPLNPNFCFLQCTSAAYPLQPEDVNLRV 199
Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEP-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVPIEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIFNLGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 12
US-10-303-128-35
; Sequence 35, Application US/10303128
; Publication No. US20030157656A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,128
; CURRENT FILING DATE: 2002-11-21
;

```

```

RESULT 11
US-10-303-118-35
; Sequence 35, Application US/10303118
; Publication No. US20030157655A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,118
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-118-35

Query Match 17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

Qy 21 IIAEIGNHOGDLDVAKRMIRMAKCGADCAKFKQSELEFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHNGSLELAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIPGNAKIS--- 75
Qy 80 TYGEKRLHLEFSDHQYRELQRYAEVGIFFTAGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEIMQKCALDYKDELALKEYTEKGLVLTSTPFSRAGANRLDMGVSAFKIGSGECNN 134
Qy 140 FPYLEKTAKGRPMVSSGMQSDMTKQVQIVKPLNPNFCFLQCTSAAYPLQPEDVNLRV 199
Db 135 YPLIKHIAAFKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSALEPGEL 259
Db 195 MLELKKEP-SCMVGLSDHTTDLNACLGAVAGACVLERHFTDSMHRSGPDIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVPIEGTILTMWMLTVKVG 318
Db 254 KELLIOSEQMAIMRGNNESKAAKQEQVTIDFASFVSVSIKDKKGEVLSMDNIWVKRPG 313
Qy 319 PKAYPPEDIFNLGKKVLTVEEDDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTQLSYE 343

RESULT 12
US-10-303-128-35
; Sequence 35, Application US/10303128
; Publication No. US20030157656A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; TITLE OF INVENTION: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,128
; CURRENT FILING DATE: 2002-11-21
;

```

```
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-128-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHGHQDLDVAKRMIRMAKCEGADCAKFKSELEFFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGIFFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLEDGMVSAFKIGSGECNN 134
QY 140 PYLEKTAAGKGRPMVVISGMSQMDTKQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMHTTNYLTPHNLRNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKTKGSDHSASLEPGEL 259
Db 195 MLELKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVVEEDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 13
US-10-303-134-35
; Sequence 35, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-128-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHGHQDLDVAKRMIRMAKCEGADCAKFKSELEFFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGIFFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLEDGMVSAFKIGSGECNN 134
QY 140 PYLEKTAAGKGRPMVVISGMSQMDTKQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMHTTNYLTPHNLRNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKTKGSDHSASLEPGEL 259
Db 195 MLELKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVVEEDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 13
US-10-303-134-35
; Sequence 35, Application US/10303134
; Publication No. US20030157657A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,134
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
```

```
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
US-10-303-134-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHGHQDLDVAKRMIRMAKCEGADCAKFKSELEFFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGIFFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLEDGMVSAFKIGSGECNN 134
QY 140 PYLEKTAAGKGRPMVVISGMSQMDTKQVQIVKPLNPNFCFLOCTSAYPLOPEDVNLRV 199
Db 135 YPLIKHIAAFKKPMIVSTGMSISIKPTVKILLDNEIPFVLMHTTNYLTPHNLRNA 194
QY 200 ISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKTKGSDHSASLEPGEL 259
Db 195 MLELKKEF-SCWVGLSDHTDNLACLGAVLACVLERHFTDSMHRSGPDIVCSMDTQAL 253
QY 260 AEL-VRSVRLVERALGSPPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMDLTVKVG 318
Db 254 KELLIQEQWAIMRGNNESKAAKQEQVTIDFAFASVVSIDIKKGEVLSMDNIWVKRPG 313
QY 319 PKAYPPEDIFNLVGKVLVTVVEEDTIMEE 348
Db 314 LGGISAAEFENILGKALRIENDTQLSYE 343

RESULT 14
US-10-303-162-35
; Sequence 35, Application US/10303162
; Publication No. US20030157658A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
; TITLE OF INVENTION: Gangliosides and Ganglioside Mimics
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/10/303,162
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/09/816,028
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 35
; LENGTH: 346
; TYPE: PRP
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: sialic acid synthase from C. jejuni OH4384 (ORF 8a)
; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus
US-10-303-162-35

Query Match      17.8%; Score 336.5; DB 14; Length 346;
Best Local Similarity 29.1%; Pred. No. 4.4e-25;
Matches 96; Conservative 63; Mismatches 164; Indels 7; Gaps 4;

QY 21 IIAEIGNHGHQDLDVAKRMIRMAKCEGADCAKFKSELEFFKFNKALE-RPYTSKHSWGK 79
Db 19 VVPEIGINHGSLAKIMVDAAFSTGAKIIKHQTHIVEDMSKAAKKVIFGNAKIS--- 75
QY 80 TYGEHKHLEFSDQYRELQRYAEVGIFFFTASGMDMAVEFLHNLNVPFFKVGSGDTNN 139
Db 76 -IYEMQKCALDYKDELALKEYTEKGLVYLSTPFSRAGANRLEDGMVSAFKIGSGECNN 134
```

Search completed: September 13, 2004, 14:55:33
Job time : 132 secs

```
Qy 140 PPYLEKTAAGRPVMISSGMSQMTMKQVQIVKPLNPNFCFLQCTSAYPELPQEDVNLRV 199
Db 135 YPLIKHTIAAFKPMIVSTGMNIESIKPTVKILLDNEIPIFVIMHTNLNLYPTPHNLVRLNA 194
Qy 200 ISEYQKLFDPDIPICYSGHETGIAISVAALGACVLERHITLTKTWKSGDHSASLEPGEL 259
Db 195 MLELKKEF-SQWVGLSDHTDNLACLAGAVALGACVLERHFTDSMRSGPDIIVCSMDTQAL 253
Qy 260 AEL-VRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILTMMDMLTVKVG 318
Db 254 KELIQSEQMAIMGNEKSKAAQEQVTIDFASFASVVSINKDKKGEVLMDNIWVKRPG 313
Qy 319 PKAYPPEDIFNLGVKKVLVTVEEDTIMEE 348
Db 314 LGGISAAEFENILGKKALRDIENDTOLSYE 343
```

```
RESULT 15
US-09-882-227-490
; Sequence 490, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Ooomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-882-227-490
```

Query Match 16.1%; Score 303; DB 10; Length 340;
Best Local Similarity 27.7%; Pred. No. 1.1e-21;
Matches 91; Conservative 66; Mismatches 149; Indels 22; Gaps 10;

```
Qy 18 PCFIIAEIGQNHQGDLDVAKRMIRMAKECGADCAKFOK-----SELEKFNKALEREPT 72
Db 4 PPKIVAEISAHNHQDNLNAKESLHAIKSGADFVKLQTYTPTSCMTLNSKEDPFIQGTLM 63
Qy 73 SKHSGKTYGSHKHLEFSPHQYRELQRYABEEVGIFFTASGMDMAVEFLHNLNVPFEKV 132
Db 64 DKENLYELYQXASTPLEW-HAELFELARKL-DLGIF--SSPSSQALELLESNCMPYKI 119
Qy 133 GSGDTNNPFYLEKTAAGRPVMISSGMSQMTMKQVQIVKPLNPNF--CFLOCTSAYPEL 190
Db 120 ASFEIVDLDLIEKAARTQKPIILSSGIATHTLQDAISLCRRVN-NFDITLLKCVSAYPS 178
Qy 191 QPEDVNLRVISEYQKLFDPDIPIGYSGHETGTATISVAALGACVLERHITLTKTWKSGDH 250
Db 179 KIEDANLLSMVKLGHPF-GVKFGLSDHTIGSLCPILATTLGASMTIEKHPILNKSLOTDS 237
Qy 251 SASLEPGELAEIVRVLVERALGSPTKQLLPCMACNEKLGKSVVAKVPIEGTILT-- 308
Db 238 AFSMDFNGFKSNVEAIKQSVLALGEEPRINPKTLEKRRFFARSILFVKIDIQKEALTEN 297
Qy 309 -MDMLTVKVG-EPKAYPPEDIFNLVGKK 334
Db 298 NIKALRPNLGLHLPKFKYK-----EILGQK 320
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:23:23 ; Search time 126 Seconds
(without alignments)
805.036 Million cell updates/sec

Title: US-10-759-277-4
Perfect score: 1887
Sequence: 1 MPLELELCPRWVGQHPGF.....EEDDTIMEELVDNHGKKIKS 359

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1887	100.0	359	3	AY96101	Aay96101 Human sia
2	1887	100.0	359	3	AY90352	Aay90352 Human gly
3	1887	100.0	359	4	AAB84683	Aab84683 Amino aci
4	1887	100.0	359	6	AAO26546	AAO26546 Human SA
5	1883	99.8	359	5	ABB90221	Abb90221 Human pol
6	1880	99.6	359	4	AAm39986	Aam39986 Human pol
7	1876	99.4	359	4	AAB93183	Aab93183 Human pro
8	1789	94.8	400	4	AAm41172	Aam41172 Human pol
9	783	41.5	338	4	ABB71364	Abb71364 Drosophil
10	544.5	28.9	338	3	AY68965	Ay68965 Cps2p pro
11	532.5	28.2	341	5	ABP26810	Abp26810 Streptoco
12	510.5	27.1	346	3	AAV96102	Aav96102 Escherich
13	510.5	27.1	346	4	AAB84684	Aab84684 Amino aci
14	510.5	27.1	346	6	AAO26547	AAO26547 Bacterial
15	419.5	22.2	125	4	ABG17251	Abg17251 Novel hum
16	395	20.9	153	4	ABG27650	Abg27650 Novel hum
17	382.5	20.3	349	5	AAm49716	Aam49716 N. mening
18	381	20.2	123	4	AAV76152	Aav76152 Human col
19	336.5	17.8	346	6	ABJ18497	Abj18497 Campyloba
20	336.5	17.8	346	6	ABJ18497	Abj18497 Campyloba
21	320.5	17.0	328	5	AAw78389	Aaw78389 Amino aci
22	303	16.1	340	2	AAw98812	Aaw98812 H. pylori
23	266.5	14.1	136	4	ABG27649	Abg27649 Novel hum
24	159.5	8.5	201	5	ABU51188	Abu51188 Helicobac
25	149.5	7.9	182	5	ABU50662	Abu50662 Helicobac

26	113.5	6.0	333	6	ABU29572	Abu29572 Protein e
27	113.5	6.0	343	7	ADC97084	Adc97084 E. faeciu
28	110	5.8	265	4	AAB96139	Aab96139 Putative
29	108.5	5.7	633	6	ABU35582	Abu35582 Protein e
30	107.5	5.7	337	6	ABU23592	Abu23592 Protein e
31	107	5.7	330	6	ABU39673	Abu39673 Protein e
32	104.5	5.5	769	6	ABM70636	Abm70636 Phototab
33	102.5	5.4	610	6	ADB06924	Adb06924 Alloiococ
34	102.5	5.4	629	6	ADB06926	Adb06926 Alloiococ
35	102.5	5.4	630	6	ADB06928	Adb06928 Alloiococ
36	101.5	5.4	337	6	ABU25380	Abu25380 Protein e
37	101.5	5.4	341	6	ABU29304	Abu29304 Protein e
38	101	5.4	66	7	ADC79295	Adc79295 Z. elonga
39	101	5.4	350	7	ADC59279	Adc59279 Comamonas
40	101	5.4	350	7	ADC52503	Adc52503 Testoster
41	100	5.3	65	7	ADC79293	Adc79293 Z. elonga
42	100	5.3	547	6	ADA12178	Ada12178 Acinetoba
43	100	5.3	547	7	ADC61133	Adc61133 Baeyer-Vi
44	100	5.3	547	7	AAE39407	Aae39407 Acinetoba
45	100	5.3	1109	6	ABU71017	Abu71017 Human adi

ALIGNMENTS

RESULT 1
AY96101
ID AY96101 standard; protein; 359 AA.

XX AC AAY96101;
XX DT 19-DEC-2000 (first entry)
XX DE Human sialic acid synthetase.
XX KW Sialic acid synthetase; human; sas gene; sialylation; glycoprotein;
XX KW plasminogen; transferrin; thyrotropin; Na+,K+-ATPase.
XX OS Homo sapiens.
XX PN WO200052135-A2.
XX PD 08-SEP-2000.
XX PF 01-MAR-2000; 2000WO-US005313.
XX PR 02-MAR-1999; 99US-0122582P.
XX PR 08-DEC-1999; 99US-0169624P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (UYJO) UNIV JOHNS HOPKINS.
XX (UYWY-) UNIV WYOMING.
XX Betenbaugh MJ, Lawrence S, Lee YC, Jarvis D, Coleman TA;
XX WPI; 2000-572178/53.
XX N-PsDB; AAAS0569.

Recombinant production of sialylated glycoproteins using cells in which the expression of enzymes, e.g. sialic acid synthetase, involved in the sialylation reaction has been altered.
Claim 16; Page 105-106; 144pp; English.
The present sequence is that of human sialic acid synthetase (SAS), an enzyme that condenses ManNAc-6-P or Man-6-P with UDP to form Neu5Ac and XDN phosphates, respectively. The sequence was deduced from SAS cDNA (see AAAS0569). Northern blots indicated ubiquitous transcription of the SAS gene in a selection of tissues. The invention provides methods and recombinantly engineered cells for producing glycoproteins having sialylated oligosaccharides. The methods involve altering the expression of enzymes involved in carbohydrate processing. A claimed cell producing sialylated glycoprotein at above endogenous levels expresses at least 1

CC (preferably human) enzyme selected from GlcNAc-2 epimerase, an enzyme
 CC catalyzing the conversion of UDP-GlcNAc to MannNAc, sialic acid
 CC synthetase, aldolase, CMP-sialic acid synthetase and CMP-sialic acid
 CC transporter at above endogenous levels. Endogenous N-
 CC acetylglucosaminidase activity may be suppressed. A claimed method for
 CC manipulating glycoprotein in an insect cell comprises enhancing the
 CC expression of 1 of the above enzymes, and a claimed method for producing
 CC sialylated glycoproteins involves expressing a heterologous protein
 CC (especially plasminogen, transferrin, Na⁺ K⁺ ATPase or thyrotropin) in
 CC the insect cell. Yeast, insect, fungal, plant and bacterial host cells
 CC can be engineered to produce new forms of sialylated glycoproteins,
 CC higher concentrations of sialylated glycoproteins and/or elevated
 CC concentrations of donor substrates (e.g. nucleotide sugars) required for
 CC sialylation
 XX
 XX
 SQ Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 3; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 Db 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTYSKHSWGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTYSKHSWGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 QY 121 FLHELNVPPFKVSGDNTNPPYLEKTAKKGRPMWISSQMSQMDTMKQYQIVKLPNPNFC 180
 Db 121 FLHELNVPPFKVSGDNTNPPYLEKTAKKGRPMWISSQMSQMDTMKQYQIVKLPNPNFC 180
 QY 181 FLQCTSAYPELQEDVNLRLVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELQEDVNLRLVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCENACNEKLGKSVVAVK 300
 Db 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCENACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359

RESULT 2
 AAY90352
 ID AAY90352 standard; protein; 359 AA.

XX AAY90352;

XX 04-DEC-2000 (first entry)

DE Human glycosylation enzyme clone HASAA37 protein sequence.

XX Human; glycosylation enzyme; glycolysis; myoglobinuria; tumour marker;
 KW immunotherapy; cosmetic surgery; metabolism; immune system disorder;
 KW haematopoietic cell deficiency; blood coagulation disorder; asthma;
 KW afibrinogenemia; blood platelet disorder; thrombocytopenia; purpura;
 KW autoimmune disorder; Addison's disease; multiple sclerosis; purpura;
 KW allergic encephalomyelitis; allergic reaction; organ rejection;
 KW graft-versus-host disease; inflammation; hyperproliferative disorder;
 KW sarcoidosis; infection; gene therapy; CMP sialic acid synthetase.

XX Homo sapiens.

OS WO200052136-A2.

XX 08-SEP-2000.

XX 01-MAR-2000; 2000WO-US0005325.

PR 02-MAR-1999; 99US-0122409P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Coleman TA;
 XX WPI; 2000-5721179/53.
 XX N-PSDB; AAA37763.
 DR New human glycosylation enzymes cytidine 5'-monophosphate sialic acid
 XX synthetase, sialic acid synthetase and aldolase and nucleic acids
 PT encoding the proteins for treating e.g., immune system disorders,
 PT microbial diseases.
 XX
 XX Claim 12; Page 110-111; 115pp; English.

CC This sequence represents a human glycosylation enzyme clone of the
 CC invention, designated HASAA37. This protein clone is a sialic acid
 CC synthetase. The sequences are useful as reagents for the differential
 CC identification of the tissues or cell types present in a biological
 CC sample, as immunological probes, for treating a disease or condition
 CC resulting from under expression of such polypeptide, for the detection
 CC and/or treatment of disorders involving aberrant glycolysis, e.g. cramps,
 CC myoglobinuria, and as tumour marker and/or immunotherapy targets. They
 CC may also be used to differentiate, proliferate and attract cells leading
 CC to the regeneration of tissues, to modulate mammalian characteristics
 CC (e.g. in cosmetic surgery) or mammalian metabolism affecting catabolism,
 CC anabolism processing, utilisation and energy storage, to change a
 CC mammal's mental state by influencing and as a food additive or
 CC preservative. The proteins can be used to assay protein levels in a
 CC sample, as a marker or detector of an immune system disorder, to inhibit
 CC cytokine activity, and as a vaccine. They may further be used to treat
 CC immune system or of haematopoietic cell deficiencies or disorders, blood
 CC coagulation disorders (e.g. afibrinogenemia), blood platelet disorders
 CC (e.g. thrombocytopenia), wounds resulting from trauma or surgery,
 CC autoimmune disorders (e.g. Addison's disease, multiple sclerosis,
 CC allergic encephalomyelitis), allergic reactions (e.g. asthma), organ
 CC rejection, graft-versus-host disease, inflammation, hyperproliferative
 CC disorders (e.g. neoplasia, purpura, sarcoidosis), diseases caused by
 CC viruses (e.g. hepatitis, meningitis, AIDS), bacteria and fungi (include
 CC e.g. tuberculosis, conjunctivitis, sepsis, typhoid, chlamydia,
 CC cellulitis), and diseases caused by parasites (e.g. amoebiasis,
 CC coccidiosis, leishmaniasis, scabies, malaria, toxoplasmosis)
 XX

SQ Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 3; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 Db 1 MPELELCGRWVGQHPFCFIIAETGQNHQGLDVAKRMIRMAKCGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTYSKHSWGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTYSKHSWGKTYGEHKHLEFSDQYRELQRYAEVEVGIFFTAGMDMAVE 120
 QY 121 FLHELNVPPFKVSGDNTNPPYLEKTAKKGRPMWISSQMSQMDTMKQYQIVKLPNPNFC 180
 Db 121 FLHELNVPPFKVSGDNTNPPYLEKTAKKGRPMWISSQMSQMDTMKQYQIVKLPNPNFC 180
 QY 181 FLQCTSAYPELQEDVNLRLVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAYPELQEDVNLRLVISEYOKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCENACNEKLGKSVVAVK 300
 Db 241 LDKTWKGSDHSASLEPGELAEVLRSVRLVERALGSPTKQLLPCENACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDIFNLVGKKVLTVVEEDDTIMEELVDNHGKKIKS 359

RESULT 3	Db	1	MPLLELCPGRWVGQHPCEIIAEIQGNHQGLDVAKEMRIRMAKCECGADCAKFOKSELEF	60
AAB84683	Qy	61	KFNKALERYTTSKHSWKTYGEBKRLHLPFSDQYRELQRYAEBVGIFFTASGMDENAVE	120
ID AAB84683 standard; protein; 359 AA.	Db	61	KFNKALERYTTSKHSWKTYGEBKRLHLPFSDQYRELQRYAEBVGIFFTASGMDENAVE	120
AC AAB84683;	Qy	121	FLHELNVPPFKVSGDNNFPYLEKTAKKGRPMVYSSGMSQMDTMKVQYQIVKPLNPNFC	180
XX 17-SEP-2001 (first entry)	Db	121	FLHELNVPPFKVSGDNNFPYLEKTAKKGRPMVYSSGMSQMDTMKVQYQIVKPLNPNFC	180
XX Amino acid sequence of a human sialic acid synthetase.	Qy	181	FLOCTSAYPLOPEDVNLRVISEYQKLPDPDIPYGYSGHETGIAISVAVALGAKVLERHIT	240
XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;	Db	181	FLOCTSAYPLOPEDVNLRVISEYQKLPDPDIPYGYSGHETGIAISVAVALGAKVLERHIT	240
KW cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;	Qy	241	LDKTWGSDHSASLEPGELAEVLVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK	300
KW sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;	Db	241	LDKTWGSDHSASLEPGELAEVLVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAKVK	300
XX vaccine.	Qy	301	IPEGTILTMDMLTVMKVGEPKAYPPEDIFNLVGGKVLVVEEDDTIMEELVDNHGKKIKS	359
XX Homo sapiens.	Db	301	IPEGTILTMDMLTVMKVGEPKAYPPEDIFNLVGGKVLVVEEDDTIMEELVDNHGKKIKS	359
XX WO200142492-A1.	RESULT 4			
XX 14-JUN-2001.	AAO26546			
XX 07-DEC-2000; 2000WO-US033136.	ID AAO26546 standard; protein; 359 AA.			
XX 09-DEC-1999; 99US-0169839P.	XX AC			
(HUMA-) HUMAN GENOME SCI INC.	XX AC			
(UYJO) UNIV JOHNS HOPKINS.	XX AC			
(UTEM) UNIV TEMPLE.	DT 06-MAR-2003 (first entry)			
(UYWY-) UNIV WYOMING.	XX			
XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;	DE Human SA synthetase protein.			
PI Jarvis D;	XX			
XX	KW Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;			
XX	KW GlcNAc-2; epimerase; UDP-GlcNAc; mannose; [Man]NAC; sialic acid;			
DR WPI; 2001-441575/47.	KW synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;			
DR N-PSDB; AAH28458.	KW transporter; sialylated glycoprotein; human.			
XX Cells producing cytidine monophosphate-sialic acid and sialylated	XX			
PT glycoprotein above endogenous levels for production of vaccines and	OS Homo sapiens.			
PT therapeutics.	XX			
XX	XX			
PS Claim 16; Fig 32; 182pp; English.	PN US2002142386-A1.			
XX	PD 03-OCT-2002.			
CC The specification describes a method for manipulating carbohydrate	XX			
CC processing pathways in cells of interest. The methods are used to	PF 16-AUG-2001; 2001US-00930440.			
CC manipulate multiple pathways involved with the sialylation reaction by	XX			
CC using recombinant DNA technology and substrate feeding approaches to	PR 02-MAR-1999; 99US-0122582P.			
CC enable the production of sialylated glycoproteins in the cells. The	PR 08-DEC-1999; 99US-0169624P.			
CC sialylation process involves the post-translational addition of the donor	PR 25-AUG-2000; 2000US-0227579P.			
CC substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific	XX			
CC acceptor carbohydrate. The cells express at least one enzyme, selected	PA (BETE/) BETENBAUGH M J.			
CC from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,	PA (LAWR/) LAWRENCE S.			
CC CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The	PA (LEEY/) LEE Y C.			
CC cells are useful for producing complex sialylated glycoproteins in cells	PA (COLE/) COLEMAN T A.			
CC of interest, especially insect cells. Glycoproteins containing sialylated	XX			
CC oligosaccharides are useful as vaccines, therapeutics and diagnostic	PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;			
CC tools. Cells producing complex sialylated glycoproteins are useful for	XX			
CC enhancing the value of heterologous expression systems and increasing the	DR N-PSDB; AAL53993.			
CC application of heterologous cell expression products as vaccines,	XX			
CC therapeutics and diagnostic tools as well as increasing the variety of	PT Manipulating glycoprotein production in insect cell, involves enhancing			
CC heterologous proteins that can be produced and lowering biotechnology	PT expression of enzymes involved in carbohydrate processing pathway such as			
CC production costs. The present sequence represents a human sialic acid	PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.			
CC synthetase, which is used in the method of the invention	XX			
XX	PS Claim 16; Fig 32; 88pp; English.			
SQ Sequence 359 AA;	XX			
Query Match	CC The invention relates to a novel method for manipulating glycoprotein			
Best Local Similarity 100.0%; Score 1887; DB 4; Length 359;	CC production in an insect cell comprising enhancing expression of an			
Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one			
	CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid			
	CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)			
	CC synthetase or CMP-SA transporter, where the expression of each enzyme is			
	CC enhanced to above endogenous levels. The novel method is useful for			
Qy	1	MPLLELCPGRWVGQHPCEIIAEIQGNHQGLDVAKEMRIRMAKCECGADCAKFOKSELEF	60	

CC manipulating glycoprotein production in an insect cell. Further methods
 CC of the invention are useful for producing sialylated glycoprotein. The
 CC sialylated glycoprotein produced by the above mentioned methods are
 CC useful as pharmaceutical compositions, vaccines, diagnostics and
 CC therapeutics. This sequence represents the human SA synthetase protein of
 CC the invention
 XX
 XX
 SQ Sequence 359 AA;

Query Match 100.0%; Score 1887; DB 6; Length 359;
 Best Local Similarity 100.0%; Pred. No. 9.4e-184;
 Matches 359; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKEGADCAKFKQSELEF 60
 Db 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKEGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTYSKHSWGKTYGEHKLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTYSKHSWGKTYGEHKLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
 QY 121 FLHELNVPPFKVSGDNTNFPVLEKTAKGRPMWISSGQSDMTMKQVQIVKPLNPFC 180
 Db 121 FLHELNVPPFKVSGDNTNFPVLEKTAKGRPMWISSGQSDMTMKQVQIVKPLNPFC 180
 QY 181 FLQCTSAVLPQEDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAVLPQEDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGDHSDASLEPGELAEIVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAVK 300
 Db 241 LDKTWKGDHSDASLEPGELAEIVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 5
 ABB90221
 ID ABB90221 standard; protein; 359 AA.
 XX
 AC ABB90221;
 XX
 DT 24-MAY-2002 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2597.
 XX
 KW Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US016450.
 XX
 PR 19-MAY-2000; 2000US-0205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-122018/16.
 DR N-PSDB; ABL90630.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 prevention of neural, immune system, muscular, reproductive,

PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX

Claim 11; SEQ ID NO 2597; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 359 AA;

Query Match 99.8%; Score 1883; DB 5; Length 359;
 Best Local Similarity 99.7%; Pred. No. 2.4e-183;
 Matches 358; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKEGADCAKFKQSELEF 60
 Db 1 MPELELCPRWVGQHPFCFIIAIGQNHQGDLDVAKRMIRWAKEGADCAKFKQSELEF 60
 QY 61 KFNKALERPPTYSKHSWGKTYGEHKLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
 Db 61 KFNKALERPPTYSKHSWGKTYGEHKLHLEFSDQYRELQRYAEVGIFFTAGMDMAVE 120
 QY 121 FLHELNVPPFKVSGDNTNFPVLEKTAKGRPMWISSGQSDMTMKQVQIVKPLNPFC 180
 Db 121 FLHELNVPPFKVSGDNTNFPVLEKTAKGRPMWISSGQSDMTMKQVQIVKPLNPFC 180
 QY 181 FLQCTSAVLPQEDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 Db 181 FLQCTSAVLPQEDVNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHIT 240
 QY 241 LDKTWKGDHSDASLEPGELAEIVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAVK 300
 Db 241 LDKTWKGDHSDASLEPGELAEIVRSVRLVERALGSPKQLLPCEMACNEKLGKSVVAVK 300
 QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 Db 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 6
 AAM39986
 ID AAM39986 standard; protein; 359 AA.
 XX
 AC AAM39986;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 3131.

XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX

QY 1 MPELELCFGRWVGQHPFCFIIAIEIGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60
Db 1 MPELELCFGRWVGQHPFCFIIAIEIGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60
QY 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE 120
Db 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE 120
QY 121 FLHELNVPPFKVSGDNTNPPYLEKTKAGGRPMVSSGQSMQMDTMKVQVIVKPLNPFC 180
Db 121 FLHELNVPPFKVSGDNTNPPYLEKTKAGGRPMVSSGQSMQMDTMKVQVIVKPLNPFC 180
QY 181 FLQCTSAYPELQEDVNLRVISYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHIT 240
Db 181 FLQCTSAYPELQEDVNLRVISYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHIT 240
QY 241 LDKTWKSDHSASLEPGEAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
Db 241 LDKTWKSDHSASLEPGEAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVK 300
QY 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
Db 301 IPEGTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359

RESULT 8
AAM41772

ID AAM41772 standard; protein; 400 AA.

XX AC AAM41772;

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 6703.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US034263.

XX PR 23-DEC-1999; 99US-00471275.

XX PR 21-JAN-2000; 2000US-00488725.

XX PR 25-APR-2000; 2000US-00552317.

XX PR 20-JUN-2000; 2000US-00598042.

XX PR 19-JUL-2000; 2000US-00620312.

XX PR 03-AUG-2000; 2000US-00653450.

XX PR 14-SEP-2000; 2000US-00662191.

XX PR 19-OCT-2000; 2000US-00693036.

XX PR 29-NOV-2000; 2000US-00727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

XX PI Zhou P, Goodrich R, Drmanac RT;

XX XX WPI; 2001-442253/47.

XX DR N-PSDB; AAI60928.

XX XX Novel nucleic acids and polypeptides, useful for treating disorders such

XX PT as central nervous system injuries.

XX PT Example 2; SEQ ID NO 6703; 10078pp; English.

XX PS

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification

XX SQ Sequence 400 AA;

Query Match 94.8%; Score 1789; DB 4; Length 400;
Best Local Similarity 88.3%; Pred. No. 1.2e-173;

Matches 346; Conservative 0; Mismatches 2; Indels 44; Gaps 1;

QY 1 MPELELCFGRWVGQHPFCFIIAIEIGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 60

Db 9 MPELELCFGRWVGQHPFCFIIAIEIGQNHQGDLDVAKRMIRMAKEGADCAKFKQSELEF 68

QY 61 KFNKALERPPTYSKSHGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE 120

Db 69 KFNKALERPPTYSKSHGKTYGEHKHLEFSDHQYRELQRYAEVGIFFTASGMDMAVE 128

QY 121 FLHELNVPPFKVSGDNTNPPYLEKTKA----- 148

Db 129 FLHELNVPPFKVSGDNTNPPYLEKTKAGHWSVLVDVCGVQLNDETSWDVLRVTS 198

QY 149 -----KGRPMVSSGQSMQMDTMKVQVIVKPLNPFCFLQCTSAYPELQEDVN 196

Db 189 KKKVLMVLVDYSGRPMVSSGQSMQMDTMKVQVIVKPLNPFCFLQCTSAYPELQEDVN 248

QY 197 LRVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHITLDTWKSGDHSASLEP 256

Db 249 LRVISEYQKLFDPDIPIGYSGHETGTAISVAALGAKVLERHITLDTWKSGDHSASLEP 308

QY 257 GELAEIVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVKIPGTLTMDMLTVKV 316

Db 309 GELAEIVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAVKIPGTLTMDMLTVKV 368

QY 317 GEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEE 348

Db 369 GEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEE 400

RESULT 9

ABBY71364

ID ABBY71364 standard; protein; 338 AA.

XX AC ABBY71364;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 40884.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX XX WO200171042-A2.

XX PD 27-SEP-2001.

XX XX 23-MAR-2001; 2001WO-US009231.

XX XX

```
PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW,
XX
XX WPI; 2001-656860/75.
XX N-PSDB; ABL15467.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 40884; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX ABBS72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 338 AA;
XX
XX Query March 41.5%; Score 783; DB 4; Length 338;
XX Best Local Similarity 45.6%; Pred. No. 6.5e-71;
XX Matches 154; Conservative 60; Mismatches 100; Indels 24; Gaps 3;
XX
XX 39 MIRMAKCGACAKFQKSELEFNKRALERPYSKHSWGKTYGHEHLEFSDHQYREL 98
XX 1 MIWEAKAGACHCVKFSKDLPAKTRSDREYISDHAWKTYGHEHLEFSDQYQL 60
XX
XX 99 QRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGDNNPPYLEKTAKKGRPMVSSG 158
XX 61 QARCKELNVDTASAMDRSLFSLNVPFKIGSGDANNFPLKKAANLNLPLVISTG 120
XX
XX 159 MQSDMTWKQYQIVKPL-NPNFLOCTSAYPELOPEDVNLRAVISEYQKLFDPIDIGYSGH 217
XX 121 MQTQTVVERIVQTRWESGKDYALMHCVSSYPTDPKCSLQLSVLRFPFNVAIGYSGH 180
XX
XX 218 ETGTAISVAAVALGAKVLERHITLDTWKGDHSGASLSPGELAEIVRSV----- 266
XX 181 ELGVIISOAVLLGARIVERHFTLDKXSKQSDHRCSLSPQELKALTITAITNFKLSSVPM 240
XX
XX 267 --RLVERALGS-----PTKQLLPCEMACNEKLGKSVVAKVPIEGITILTMDLTV 314
XX 241 PQEIVTKLNGDEEALAQHVESKTIPLCELPFCRNKLGKSIVAARNLNGYRLQADMAI 300
XX
XX 315 KVGEPKAYPPEDIENLVGKVLVTVEEDDTIMEELVDN 352
XX 301 KVSEFSGLTABDFDLVGLKELADNIGEDPEILNGSIIN 338
XX
XX RESULT 10
XX AAY68965.
XX ID AAY68965 standard; protein; 338 AA.
XX AC
XX AAY68965;
XX
XX 30-MAY-2000 (first entry)
XX
XX Cps2P protein which is involved in stalic acid synthesis.
XX
XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;
XX capsular component; antigen; regulation; chain length determination;
XX complement-mediated opsonophagocytosis; serotype-specific detection;
XX antigen; vaccine; Streptococcal disease; ORF 2Z; ORF 2Y; ORF 2Z; Cps2A;
XX Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J; Cps2K;
```

```
XX Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.
XX Streptococcus suis.
XX WO200005378-A2.
XX
XX 03-FEB-2000.
XX
XX 19-JUL-1999; 99WO-NL000460.
XX
XX 22-JUL-1998; 98EP-00202465.
XX 22-JUL-1998; 98EP-00202467.
XX
XX (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX
XX Smith HE;
XX
XX WPI; 2000-195104/17.
XX N-PSDB; AAZ60929.
XX
XX New nucleic acid containing the capsular gene cluster of Streptococcus
XX suis, used for serotype-specific detection and to generate antigens or
XX mutants for vaccination.
XX
XX Disclosure; Fig 3; 144pp; English.
XX
XX The proteins AAY68950-69 are encoded by the capsular gene cluster of
XX Streptococcus suis serotype 2. The genes in this cluster are involved in
XX polysaccharide biosynthesis of capsular components and antigens. The
XX proteins are involved in regulation (CpsA), chain length determination
XX (CpsB, CpsC), export (CpsC), and biosynthesis (CpsE, CpsF, CpsG, CpsH,
XX CpsI, CpsK). The capsule confers bacterium resistance to complement-
XX mediated opsonophagocytosis. The gene cluster is used as a source of
XX probes and primers for serotype-specific detection of S. suis and is also
XX useful for recombinant production of the proteins. The proteins are then
XX useful for producing antigens that can be used in vaccines, for
XX controlling or eradicating a Streptococcal disease, in humans or animals,
XX e.g. against S. suis in pigs
XX
XX Sequence 338 AA;
XX
XX Query Match 28.9%; Score 544.5; DB 3; Length 338;
XX Best Local Similarity 36.0%; Pred. No. 1.6e-46;
XX Matches 123; Conservative 73; Mismatches 129; Indels 17; Gaps 8;
XX
XX 20 FIIEIGQNHGQDLDVAKRMIRMAKCEGACAKFQKSELEFNKRALERPYSKHSWGK 79
XX 3 YIIAIEGNCNHGQDVLARKMVEVAVDCGVDAVKFTFKADLLISKYAPKAEY-QKITTGE 61
XX
XX 80 TYG--EHKHELFSDHQYRELQRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGSDT 137
XX 62 SDSQLEMTRRLELPEEYLDLDYCLEKGVDFVSTPFDERSLDFLISDTMPYKIPSEI 121
XX
XX 138 NNFPYLEKTAKKGRPMVSSGMSQMDTMKQYQIVKPLNPN----FCFLOCTSAYPELOPE 193
XX 122 TNLPLEKIGQAKKVLSTGMAVND---EIHQAVKIIQENGTTDISILHCTTEYTPVP 178
XX
XX 194 DVNLRVISEYQKLFDPIDIGYSGHETGTIAISVAAVALGAKVLERHITLDTWKGDHSGAS 253
XX 179 ALNLNLVTLTKKEFENLTIGYSDHSGSEVPITAAAMGAELIEKHFTLIDNEMEGPDHKS 238
XX
XX 254 LEPGELAEIVRSVRLVERALGSPKQLLPCEMACNEKL--GKSVVAKVPIEGITILTM 311
XX 239 ATPDILAAALVKGVRIVEQSLGKFEK--PEEVEVRNKIVARKSVAKIAKGEVFTTEEN 296
XX
XX 312 LTVKVGEP-KAYPPEDIENLVGKVLVTVEEDDTIMEELVDN 352
XX 297 ITVK--RPGNGISPMWYKVLGVQSEQDFEEDQNICHSAFEN 336
XX
XX RESULT 11
XX ABP26810
XX ID ABP26810 standard; protein; 341 AA.
```


XX SQ Sequence 346 AA;

Query Match 27.1%; Score 510.5; DB 3; Length 346;

Best Local Similarity 36.3%; Pred. No. 4.9e-43;

Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FIIAIGQNHQGLDVAKRMRMAKCGADCAKFKQSELEFFKFNKALERPYSKHSWG- 78

DB 5 YIVAEIGCNHNGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPKAEYQIKNTGEL 64

QY 79 KTYGEHRHLEFSDHQRELQRYAEVGIPTFASGMDMAVEFLHNLVFFKVGSGDTN 138

DB 65 ESQLEMTKLEMYDDYLHLMEYAVSLNLDVFTFDFEDSIDFLASLKQKIWKIPSGELL 124

QY 139 NFPVLEKTAKKGRP---WVSSGQSMQMDTMKVQVQIV---KPLNPNFCFLQCTSAYPELP 192

DB 125 NLPVLEKIAKLPIPDKKIISTGMATIDEIKQSVSIFINNKPVGNTILHCNTEYPTPF 184

QY 193 EDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSA 252

DB 185 EDVNLNAINDLKHFPPKNNIGFSDHSSGFYAAIAAAPPYGITFIEKHFTLDKSMGPDHLA 244

QY 253 SLEPGEIAELVRSVRLVERALGSPKQLPCMACNEKLGKSVVAKVPIPGTILTMDEL 312

DB 245 SIEPDELKHLGICVRCVEKSLGNSKVVTASERKNKIVARKSIIAKTEIKKGEVFEKNI 304

QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVEEDDTIMEELV 350

DB 305 TTK--RPGNGISPWEWYLLGK-----IAEQDFIPDELI 336

RESULT 13

AAB84684

ID AAB84684 standard; protein; 346 AA.

AC AAB84684;

XX 17-SEP-2001 (first entry)

XX Amino acid sequence of a bacterial sialic acid synthetase.

XX Carbohydrate processing; sialylation; sialylated glycoprotein; CMP-SA;

XX cytidine monophosphate-sialic acid; N-acetylglucosamine-2 epimerase;

XX sialic acid synthetase; aldolase; CMP-SA synthetase; CMP-SA transporter;

XX vaccine.

XX Escherichia coli.

XX WO200142492-A1.

XX 14-JUN-2001.

XX 07-DEC-2000; 2000WO-US033136.

XX 09-DEC-1999; 99US-0169839P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (UYJO) UNIV JOHNS HOPKINS.

XX (UTEM) UNIV TEMPLE.

XX (UYWY-) UNIV WYOMING.

XX Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA, Palter K;

XX Jarvis D;

XX WPI; 2001-441575/47.

XX DR N-PSDB; AAH28459.

XX Cells producing cytidine monophosphate-sialic acid and sialylated

XX glycoprotein above endogenous levels for production of vaccines and

XX therapeutics.

XX Example 5; Page 165-166; 182pp; English.

XX The specification describes a method for manipulating carbohydrate

XX processing pathways in cells of interest. The methods are used to

XX manipulate multiple pathways involved with the sialylation reaction by

XX using recombinant DNA technology and substrate feeding approaches to

XX enable the production of sialylated glycoproteins in the cells. The

XX sialylation process involves the post-translational addition of the donor

XX substrate cytidine monophosphate-sialic acid (CMP-SA) onto a specific

XX acceptor carbohydrate. The cells express at least one enzyme, selected

XX from N-acetylglucosamine-2 epimerase, sialic acid synthetase, aldolase,

XX CMP-SA synthetase and CMP-SA transporter, above endogenous levels. The

XX cells are useful for producing complex sialylated glycoproteins in cells

XX of interest, especially insect cells. Glycoproteins containing sialylated

XX oligosaccharides are useful as vaccines, therapeutics and diagnostic

XX tools. Cells producing complex sialylated glycoproteins are useful for

XX enhancing the value of heterologous expression systems and increasing the

XX application of heterologous cell expression products as vaccines,

XX therapeutics and diagnostic tools as well as increasing the variety of

XX heterologous proteins that can be produced and lowering biotechnology

XX production costs. The present sequence represents a sialic acid

XX synthetase (neub), which is used in the method of the invention

XX SQ Sequence 346 AA;

Query Match 27.1%; Score 510.5; DB 4; Length 346;

Best Local Similarity 36.3%; Pred. No. 4.9e-43;

Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;

QY 20 FIIAIGQNHQGLDVAKRMRMAKCGADCAKFKQSELEFFKFNKALERPYSKHSWG- 78

DB 5 YIVAEIGCNHNGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPKAEYQIKNTGEL 64

QY 79 KTYGEHRHLEFSDHQRELQRYAEVGIPTFASGMDMAVEFLHNLVFFKVGSGDTN 138

DB 65 ESQLEMTKLEMYDDYLHLMEYAVSLNLDVFTFDFEDSIDFLASLKQKIWKIPSGELL 124

QY 139 NFPVLEKTAKKGRP---WVSSGQSMQMDTMKVQVQIV---KPLNPNFCFLQCTSAYPELP 192

DB 125 NLPVLEKIAKLPIPDKKIISTGMATIDEIKQSVSIFINNKPVGNTILHCNTEYPTPF 184

QY 193 EDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHITLDTKWGSDHSA 252

DB 185 EDVNLNAINDLKHFPPKNNIGFSDHSSGFYAAIAAAPPYGITFIEKHFTLDKSMGPDHLA 244

QY 253 SLEPGEIAELVRSVRLVERALGSPKQLPCMACNEKLGKSVVAKVPIPGTILTMDEL 312

DB 245 SIEPDELKHLGICVRCVEKSLGNSKVVTASERKNKIVARKSIIAKTEIKKGEVFEKNI 304

QY 313 TVKVGEP-KAYPPEDIFNLVGKVLVTVEEDDTIMEELV 350

DB 305 TTK--RPGNGISPWEWYLLGK-----IAEQDFIPDELI 336

RESULT 14

AAO26547

ID AAO26547 standard; protein; 346 AA.

AC AAO26547;

XX 06-MAR-2003 (first entry)

XX Bacterial sialic acid synthetase (NeuB) protein.

XX Vaccine; glycoprotein; insect cell; enzyme; N-acetylglucosamine-2;

XX GlcNAc-2; epimerase; UDP-GlcNAc; mannose; (Man)NAC; sialic acid;

XX synthetase; aldolase; cytidine monophosphate-sialic acid; CMP-SA;

XX transporter; sialylated glycoprotein.

XX Escherichia coli.

XX US2002142386-A1.

XX 03-OCT-2002.

```
XX PF 16-AUG-2001; 2001US-00930440.
XX PR 02-MAR-1999; 99US-0122582P.
XX PR 08-DEC-1999; 99US-0169624P.
XX PR 25-AUG-2000; 2000US-0227579P.
XX PA (BETE/) BETENBAUGH M J.
XX PA (LAWR/) LAWRENCE S.
XX PA (LEEV/) LEE Y C.
XX PA (COLE/) COLEMAN T A.
XX PI Betenbaugh MJ, Lawrence S, Lee YC, Coleman TA;
XX DR WPI; 2003-102519/09.
XX DR N-PSDB; AAL53994.
XX PT Manipulating glycoprotein production in insect cell, involves enhancing
XX PT expression of enzymes involved in carbohydrate processing pathway such as
XX PT N-acetylglucosamine-2 epimerase or sialic acid synthetase.
XX PS Disclosure; Fig 35D; 88pp; English.
XX CC The invention relates to a novel method for manipulating glycoprotein
XX CC production in an insect cell comprising enhancing expression of an
XX CC enzyme, such as N-acetylglucosamine-2 (GlcNAc-2) epimerase, one
XX CC catalysing conversion of UDP-GlcNAc to mannose (Man)NAC, sialic acid
XX CC synthetase, aldolase, cytidine monophosphate-sialic acid (CMP-SA)
XX CC synthetase or CMP-SA transporter, where the expression of each enzyme is
XX CC enhanced to above endogenous levels. The novel method is useful for
XX CC manipulating glycoprotein production in an insect cell. Further methods
XX CC of the invention are useful for producing sialylated glycoprotein. The
XX CC sialylated glycoprotein produced by the above mentioned methods are
XX CC useful as pharmaceutical compositions, vaccines, diagnostics and
XX CC therapeutics. This sequence represents the bacterial sialic acid (NeuB)
XX CC synthetase protein of the invention
XX SQ Sequence 346 AA;
XX Query Match 27.1%; Score 510.5; DB 6; Length 346;
XX Best Local Similarity 36.3%; Pred. No. 4.9e-43;
XX Matches 123; Conservative 61; Mismatches 140; Indels 15; Gaps 6;
QY 20 FTIABIGNHQDGLDVAKRMIRMAKECGADCAKFKSELEFKFNKALERPPTSKHSG- 78
Db 5 YIVABIGCNHGSVDIAREMILKAKEAGVNAVVFQTFKADKLISAIAPKAEYQIKNTGEL 64
QY 79 KTYGKHKLHPSHDQYRELQRYAEVGIFFTASGDMEMAVFLHNLVPPFKVSGSDTN 138
Db 65 ESQLEMTKKLEMKYDDYLHLMGAYSLNLDVSTFPDESDSLASLKKQIKWIPSGELL 124
QY 139 NPPYLEKTAKKGRP---MVISGMSQMDMTMKQVQIV---KPLNPNFCFLOCTSAVPLQP 192
Db 125 NLPLYELKIAKLPIPKKIIISTGMATIDBIKQSVSIFINNVKVPVGNITLHCNTYPTPP 184
QY 193 EDVNLRVITSEYQKLPDDPIGYSGHETGIAISVAVALGAKYLERHITLDTKWGSDRSA 252
Db 185 EDVNLNAINDLKKHFPKNNIGFSDHSGFYAIAAAPPYGITIEKHFTLTKSMSPDHLA 244
QY 253 SLEPELAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVAKVIPGEGTILTMOML 312
Db 245 SIEPEDELKHLICIGVRCVSEKSLGNSKSVTVASERKNKIVARKSIIATIKKGEVFSKNI 304
QY 313 TKVKGEP-KAYPPBEDIFNLVGKVLVTVEEDDTIMEELV 350
Db 305 TTK--RPGNGGISPMWYNLLGK-----IAEQDFIPDELI 336
RESULT 15
ABG17251
ID ABG17251 standard; protein; 125 AA.
XX
AC ABG17251;
```

```
XX 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #17242.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX XX 30-MAR-2001; 2001WO-US008631.
XX XX 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX XX (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS81438.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 47610; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (II) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensic, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 125 AA;
XX Query Match 22.2%; Score 419.5; DB 4; Length 125;
XX Best Local Similarity 79.2%; Pred. No. 2e-34;
XX Matches 80; Conservative 6; Mismatches 14; Indels 1; Gaps 1;
QY 1 MPLELELCFGRWVGQHPCFIIAEIGQNHQDGLDVAKRMIRMAKECGADCAKFKSELEF 60
Db 9 NPVELELCFGRWVGQHPCFIIAEIGQNHQDGLTVAKRMIRMAKECGADCAKFKSELEF 68
QY 61 KFNKALERPPTSKHSGMK-TYGEHKHLEFSDQYRELQ 100
Db 69 KFNKALDRPPTSKHSLGEDVRGAQTDIWKFNHDDVQGAAR 109
Search completed: September 13, 2004, 14:41:01
Job time : 130 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 13, 2004, 14:34:23 ; Search time 118 Seconds
(without alignments)
959.924 Million cell updates/sec

Title: US-10-759-277-4

Perfect score: 187

Sequence: 1 MPLELELCPRWVGQHPHF.....BEDDTIMEELVDNHGKKIKS 359

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mbc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriaph:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1810	95.9	359	11 Q99J77	Q99J77 mus musculu
2	1804	95.6	359	11 Q99JH0	Q99JH0 mus musculu
3	854	45.3	372	5 Q9VG74	Q9VG74 dtrosophila
4	547.5	29.0	280	16 Q8KE65	Q8KE65 chlorobium
5	534.5	28.3	341	2 Q87186	Q87186 streptococc
6	534.5	28.3	341	16 Q8E501	Q8E501 streptococc
7	532.5	28.2	339	2 Q9ALW6	Q9ALW6 streptococc
8	532.5	28.2	339	2 Q93TI2	Q93TI2 streptococc
9	532.5	28.2	341	2 Q9RPC0	Q9RPC0 streptococc
10	532.5	28.2	341	16 Q8DZ83	Q8DZ83 streptococc
11	526	27.9	749	16 Q82UC3	Q82UC3 nitrosomona
12	524	27.8	311	16 Q82HY4	Q82HY4 streptomyc
13	518	27.5	334	2 Q7X523	Q7X523 campylobact
14	516.5	27.4	333	16 Q8DDZ7	Q8DDZ7 vibrio vuln
15	514	27.2	361	2 Q8KNA2	Q8KNA2 pseudomonas
16	512	27.1	312	16 Q9AK45	Q9AK45 streptomyc

17	510.5	27.1	346	2 Q46675	Q46675 escherichia
18	510	27.0	338	2 Q9RDX5	Q9RDX5 legionella
19	509.5	27.0	351	16 Q893U6	Q893U6 clostridium
20	507.5	26.9	753	16 Q8F5Q4	Q8F5Q4 leptospira
21	506.5	26.8	341	2 Q9AQ16	Q9AQ16 streptococ
22	506.5	26.5	641	16 Q89HL9	Q89HL9 bradyrhizob
23	497	26.3	754	16 Q7U964	Q7U964 synecococc
24	487	25.8	334	16 Q9PMX2	Q9PMX2 campylobact
25	472	25.0	333	16 Q7U911	Q7U911 synecococc
26	472	25.0	346	2 Q93NQ2	Q93NQ2 escherichia
27	461.5	24.5	356	16 Q89HJ8	Q89HJ8 bradyrhizob
28	437.5	23.2	286	17 Q8TUL5	Q8TUL5 methanosarc
29	435.5	23.1	344	16 Q8FSQ8	Q8FSQ8 leptospira
30	435	23.1	346	16 Q8A711	Q8A711 bacteroides
31	431.5	22.9	357	16 Q87T70	Q87T70 vibrio para
32	419	22.2	350	16 Q97H26	Q97H26 clostridium
33	410	21.7	351	2 Q8KH52	Q8KH52 pseudomonas
34	393.5	20.9	352	2 Q9R9S2	Q9R9S2 aeromonas p
35	390.5	20.7	351	16 Q8R1B8	Q8R1B8 fusobacteri
36	384.5	20.4	351	2 Q8KN54	Q8KN54 pseudomonas
37	382.5	20.3	349	16 Q57265	Q57265 neisseria m
38	381.5	20.2	332	16 Q8F330	Q8F330 leptospira
39	374.5	19.8	354	16 Q7V953	Q7V953 prochloroco
40	367.5	19.5	343	2 Q93J08	Q93J08 campylobact
41	363.5	19.3	354	2 Q9R9M1	Q9R9M1 rhizobium m
42	360.5	19.1	229	2 Q9F9F4	Q9F9F4 campylobact
43	358	19.0	356	16 Q9A4G6	Q9A4G6 caulobacter
44	356.5	18.9	343	16 Q9PMY2	Q9PMY2 campylobact
45	356	18.9	343	2 Q9EU02	Q9EU02 campylobact

ALIGNMENTS

RESULT 1

Q99J77					
ID	Q99J77	PRELIMINARY;	PRT;	359	AA.
AC	Q99J77;				
DT	01-JUN-2001 (TREMBlrel. 17, Created)				
DT	01-JUN-2001 (TREMBlrel. 17, Last sequence update)				
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)				
DE	Similar to N-acetylneuraminic acid phosphate synthase, sialic acid synthase.				
GN	NANS OR 4632418E04RIK OR SAS.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]_TaxID=10090;				
RP	SEQUENCE FROM N.A.				
RA	Strausberg R.;				
RL	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Skin;				
RX	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."				
RL	Nature 420:563-573 (2002).				
DR	EMBL; BC003307; AAH03307.1; -.				
DR	HSSP; AK076290; BAC36290.1; -.				
DR	HSSP; P19614; 2J1A.				
DR	MGD; MGI:2149820; Nans.				
DR	GO; GO:0005829; C:cytosol; IDA.				
DR	GO; GO:0019007; F:N-acetylneuraminic acid phosphate synthase . . . ; IDA.				
DR	InterPro; IPR006013; AntifreezeII.				
DR	InterPro; IPR006014; Antifreeze-Dom.				
DR	InterPro; IPR006190; Antifreeze_like.				
DR	Pfam; PF01354; NeuB.				
DR	Pfam; PF01354; Antifreeze; 1.				
DR	Pfam; PF01302; NeuB; 1.				

DR PRINTS: PR00357: ANTIFREEZEIII.
 DR Prodom: PD003258; AntifreezeIII; 1.
 DR PROSITE: PS0844; AFP LIKE; 1.
 SQ SEQUENCE 359 AA; 40024 MW; 4666CB883558A373 CRC64;

Query Match 95.9%; Score 1810; DB 11; Length 359;
 Best Local Similarity 94.4%; Pred. No. 8.2e-141;
 Matches 339; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 MPELELCPRWVGKHPFCIIAIEIGNHOGDLDVAKRMIRTAKEGADCAKFKQSELEF 60
 |||||
 Db 1 MPELELCPRWVGKHPFCIIAIEIGNHOGDLDVAKRMIRTAKEGADCAKFKQSELEF 60
 |||||

QY 61 KFNKALERPYSKSHWGKTYGHEKHLEFSDHQYRELQRYAEVGIFFTASGMDENAVE 120
 |||||
 Db 61 KFNKALERPYSKSHWGKTYGHEKHLEFSDHQYRELQRYAEVGIFFTASGMDENAVE 120
 |||||

QY 121 FLHELNVPPFKVSGGDTNNPFPYLEKTAKKGRPMVSISSGQSMQMDTKQVQIVKPLPNFC 180
 |||||
 Db 121 FLHELNVPPFKVSGGDTNNPFPYLEKTAKKGRPMVSISSGQSMQMDTKQVQIVKPLPNFC 180
 |||||

QY 181 FLQCTSAYPEQEDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240
 |||||
 Db 181 FLQCTSAYPEQEDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240
 |||||

QY 241 LDKTWKGS DSHSASLEPGEIAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
 |||||
 Db 241 LDKTWKGS DSHSASLEPGEIAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
 |||||

QY 301 IPEGTTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 |||||
 Db 301 IPEGTTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 |||||

RESULT 2
 Q9JH0 PRELIMINARY; PRT; 359 AA.
 ID Q9JH0
 AC Q9JH0; 15, Created
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE N-acetylneuraminic acid 9-phosphate synthetase.
 GN NANS OR SAS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20334323; PubMed=10873658;
 RA Nakata D., Close B.E., Colley K.J., Matsuda T., Kitajima K.;
 RT "Molecular cloning and expression of the mouse N-acetylneuraminic acid
 RT 9-phosphate synthase which has not the deaminoneuraminic acid (XDN)
 RL Biochem. Biophys. Res. Commun. 273:642-648 (2000).
 DR EMBL: AB041263; BAA98131.1; -
 DR PIR: JC7321; JC7321.
 DR HSP: P19614; 2JIA.
 DR MGD; MGI:2149820; Nans.
 DR GO; GO:0005829; Cytosol; IDA.
 DR GO; GO:0019007; F-N-acetylneuraminic acid phosphate synthase . . ; IDA.
 DR InterPro; IPR006190; Antifreeze_like.
 DR InterPro; IPR004144; Neuf.
 DR Pfam; PF03102; NeufB; 1.
 DR PROSITE; PS0844; AFP LIKE; 1.
 SQ SEQUENCE 359 AA; 39994 MW; 5FFB7D40C558A373 CRC64;

Query Match 95.6%; Score 1804; DB 11; Length 359;
 Best Local Similarity 94.2%; Pred. No. 2.5e-140;
 Matches 338; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 MPELELCPRWVGKHPFCIIAIEIGNHOGDLDVAKRMIRTAKEGADCAKFKQSELEF 60
 |||||
 Db 1 MPELELCPRWVGKHPFCIIAIEIGNHOGDLDVAKRMIRTAKEGADCAKFKQSELEF 60
 |||||

QY 61 KFNKALERPYSKSHWGKTYGHEKHLEFSDHQYRELQRYAEVGIFFTASGMDENAVE 120
 |||||
 Db 61 KFNKALERPYSKSHWGKTYGHEKHLEFSDHQYRELQRYAEVGIFFTASGMDENAVE 120
 |||||

QY 121 FLHELNVPPFKVSGGDTNNPFPYLEKTAKKGRPMVSISSGQSMQMDTKQVQIVKPLPNFC 180
 |||||
 Db 121 FLHELNVPPFKVSGGDTNNPFPYLEKTAKKGRPMVSISSGQSMQMDTKQVQIVKPLPNFC 180
 |||||

QY 181 FLQCTSAYPEQEDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240
 |||||
 Db 181 FLQCTSAYPEQEDVNLRLVISEYQKLPDPIPIGYSGHETGIAISVAALGAKVLERHIT 240
 |||||

QY 241 LDKTWKGS DSHSASLEPGEIAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
 |||||
 Db 241 LDKTWKGS DSHSASLEPGEIAELVRSVRLVERALGSPTKQLLPCEMACNEKLGKSVVAKVK 300
 |||||

QY 301 IPEGTTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 |||||
 Db 301 IPEGTTILTMDLTVKVGEPKAYPPEDI FNLVGGKVLVTVEEDDTIMEELVDNHGKKIKS 359
 |||||

RESULT 3
 Q9VG74 PRELIMINARY; PRT; 372 AA.
 ID Q9VG74
 AC Q9VG74; Q9SVII;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE CG5232 protein (Neu5Ac synthase).
 GN NEUSAC OR CG5232.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Amanatides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kvatits S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,


```

QY 249 DHSASLEPCELAEVLRSVRLVERALGSPTRKQLLPCEWACNEKL 291
DB 235 DQASVEISGSLVSRNIRDIEKALGDGVKRVYDGEAARKKL 277

RESULT 5
O87186
ID O87186 PRELIMINARY; PRT; 341 AA.
AC O87186;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NEUB protein.
GN NEUB
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamoto S., Miyake K., Iijima S.;
RT "Identification and Characterization of cps (capsular polysaccharide)
RT Genes from Streptococcus agalactiae Type Ia.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Miyake K., Yamamoto S., Koike Y., Watanabe M., Iijima S.;
RT "Molecular Characterization of Type-Specific Capsular Polysaccharide
RT Biosynthesis Genes of Streptococcus agalactiae Type Ia.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AB017355; BAA33753.1; -.
DR EMBL; AB028896; BAA82287.1; -.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR Pfam; PF01354; NeuB; 1.
DR Pfam; PF03102; NeuB; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A340 CRC64;

Query Match 28.3%; Score 534.5; DB 2; Length 341;
Best Local Similarity 36.8%; Pred. No. 1.1e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAIEGQNHGDLVAKRMIRMAKCGADCAKFKSELEFKNRKALRPYTSKHSWGK 79
DB 3 YIIAIEGCHNGDINLAKMVDVAVSCGVDVAVKQTFKAEKLSKPAKAEY-QKETTGT 61

QY 80 TYG--EHRHLEFSDQYRELQRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGDT 137
DB 62 ADSQLEMTKRLSLSPFEYLEMRDYALSKGVETFTPTFDEESLEFLISTDMPYIKPSGEI 121

QY 138 NNFPYLEKTAAGRPVMISSGQSDMTMKQVQIVKPLNP-----FCFLQCTSAAYLQPE 193
DB 122 TNLPLYLEKIGKQKKVILSTGMA---VMEEIHQAVNLRQNGTTDISILHCTEYPTYP 178

QY 194 DYNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSAS 253
DB 179 SINLNVHTLKDFEKDLTIGYSDHSIGSEVPVIAAAMGAIEVKHFTLDTNWEGPDHKAS 238

QY 254 LEPGELAEILRSVRLVERALGSPTRKQLLPCEWACNEKLGVVAVKPIPEGTILTMDMLT 313
DB 239 ATPDILALVKGVRIVEQALGRFEKIPDPVEEKNKIVARKSVVALPKPIKKGDIYSIENT 298

QY 314 VKVGP-KAYPDEDINLVGKKVLTVEEDDTIMEELVDN 352
DB 299 VK--RPGNGISPMNWDILGQEAQDDFEDEVIRDSRFEN 336

RESULT 6
O8E501
ID O8E501 PRELIMINARY; PRT; 341 AA.
AC O8E501;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein.
GN NEUB OR GBS1236.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
MEDLINE=224242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevallier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lailou L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL768849; CAD46895.1; -.
DR Sagalish; gbs1236; -.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; NeuB.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; NeuB; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 341 AA; 38091 MW; 17620BF82A340 CRC64;

Query Match 28.3%; Score 534.5; DB 16; Length 341;
Best Local Similarity 36.8%; Pred. No. 1.1e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FIIAIEGQNHGDLVAKRMIRMAKCGADCAKFKSELEFKNRKALRPYTSKHSWGK 79
DB 3 YIIAIEGCHNGDINLAKMVDVAVSCGVDVAVKQTFKAEKLSKPAKAEY-QKETTGT 61

QY 80 TYG--EHRHLEFSDQYRELQRYAEVGIFFTAGMDMAVEFLHMLNVPFKVSGDT 137
DB 62 ADSQLEMTKRLSLSPFEYLEMRDYALSKGVETFTPTFDEESLEFLISTDMPYIKPSGEI 121

QY 138 NNFPYLEKTAAGRPVMISSGQSDMTMKQVQIVKPLNP-----FCFLQCTSAAYLQPE 193
DB 122 TNLPLYLEKIGKQKKVILSTGMA---VMEEIHQAVNLRQNGTTDISILHCTEYPTYP 178

QY 194 DYNLRVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSAS 253
DB 179 SINLNVHTLKDFEKDLTIGYSDHSIGSEVPVIAAAMGAIEVKHFTLDTNWEGPDHKAS 238

QY 254 LEPGELAEILRSVRLVERALGSPTRKQLLPCEWACNEKLGVVAVKPIPEGTILTMDMLT 313
DB 239 ATPDILALVKGVRIVEQALGRFEKIPDPVEEKNKIVARKSVVALPKPIKKGDIYSIENT 298

QY 314 VKVGP-KAYPDEDINLVGKKVLTVEEDDTIMEELVDN 352
DB 299 VK--RPGNGISPMNWDILGQEAQDDFEDEVIRDSRFEN 336

RESULT 7
O9ALW6
ID O9ALW6 PRELIMINARY; PRT; 339 AA.
AC O9ALW6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative N-acetyl neuraminic acid synthetase NeuB.
GN NEUB.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

```

```

OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=NT6;
RC McKinnon K., Chaffin D.O., Rubens C.E.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF337958; AAK11670.1; -- biosynthesis; IEA.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze like.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; Antifreeze; 1.
DR PROSITE; PS0844; AFP LIKE; 1.
SQ SEQUENCE 339 AA; 37773 MW; 156C2DEBF1D453C7 CRC64;

Query Match 28.2%; Score 532.5; DB 2; Length 339;
Best Local Similarity 36.8%; Pred. No. 1.6e-35;
Matches 125; Conservative 66; Mismatches 136; Indels 13; Gaps 6;

QY 20 FTIAEIGNHGDLDVAKEMIRMAKCEGADCAKFKSELEFKNRKALERPPTS KHSWGK 79
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 YIIAIEGCHNGDINLAKQWVAVSCGVDVAVKQTFKAEKLSKFAKPAEY-QKATTGT 61
QY 80 TYG--EHKHLEFSDHQRELORYAEVEGIFTASGMDMAVEFLHNLNVPFKVSGDT 137
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 ADSQLEMTKRLSEFEEYLEMDYAIKGVETFTFPFDESLEFLISTDMPYIKPSGEI 121
QY 138 NNFPYLEKTAKGRPMWISSGMSMDTKQVQIVKPLNPN----FCFLQCTSAVPLQPE 193
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 TNLPYLEKIGKQKQKVILTGNMA---VMEEIHQAVNLRQNGTTDISILHCTEYPTYP 178
QY 194 DYNLAVISEYQKLPDIPIGYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSAS 253
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 SUNLNVHTLKDEKDLTIGYSDHSIGSEVPFAAAMGAIEVKHFTLDTNMEGPDHKAS 238
QY 254 LEPGLAEIVRSVRLVERALGSPTQLLPCEMACNEKLGKSVVAKVPIEGHILTMDMIT 313
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 ATPDILAAVKGRIVEQALGRFEKIPDVEEKNKIVARKSVVALPKPKKGDIYSIENT 298
QY 314 VKVGP-KAYPEDIFNLVKGKVLTVREDDTIMBELVDN 352
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 VK--RPGNGISPMWYDILGQEAQDDFEDEVIRSRFEN 336

RESULT 9
Q9RPCO PRELIMINARY; PRT; 341 AA.
AC Q9RPCO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE CPSN (NEUB).
GN CPSN OR NEUB.
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=1311;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=COHI;
RX MEDLINE=93360815; PubMed=8355611;
RA Rubens C.E., Heggen L.M., Haft R.F., Wessels M.R.;
RT "Identification of cpsD, a gene essential for type III capsule expression in group B streptococci.";
RL Mol. Microbiol. 8:843-855(1993).
RN [2] SEQUENCE FROM N.A.
RP STRAIN=COHI;
RC Chaffin D.O., Yim H.H., Beres S.B., Sweet E.S., Nittayajarn A., Rubens C.E.;
RA "Capsular Polysaccharide Synthesis Locus of Streptococcus agalactiae, Serotype III.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3] SEQUENCE FROM N.A.
RP STRAIN=CNCTC 1/82;
RC McKinnon K., Chaffin D.O., Rubens C.E.;
RA "Streptococcus agalactiae type V polysaccharide synthesis operon complete sequence.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF163833; AAD53074.1; --
DR EMBL; AF349539; AAK29661.1; --
DR PIR; T44651; T44651.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze like.
DR InterPro; IPR004144; NeuB.

```


Matches 123; Conservative 62; Mismatches 135; Indels 16; Gaps 5;

```
QY 21 IIAEIGNHOGDLDVAKEMIRMAKCGADCAKFKSELEFKNRKALEPPTSK-HSWGK 79
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
6 IIAEAGVNHGDLNLAKKLLEVAASKGADPFVKFQSKAEKLVCSKNAKKAAYQLKTTAKDE 65
QY 80 TYGEHKHLEFSDHQYRELQRYAEVEGIFFTASGMDMAVEFLHNLVPPFKVSGGDTNN 139
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
66 SOLEMIKKLEDFNAHQQLLISHAKQCGIAFLSTAFDLESIELLDLGLGVFKIPSGEITN 125
QY 140 FPYLEKTAAGRPWVSSGMSQMDTMKVQYIV---KPLNPNFCFLOCTSAYPLOQPDVN 196
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
126 LPYLKKAIAKLNKIIILSTGMSNLGEIEAALVLCKEGTQRANITLLHCTTEYPAPFNEV 185
QY 197 LRIVSEYQKLFDPDIPICYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSASLEP 256
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
186 LKAWQILKNAF-NLDVGYSHTGIIHSLTAIGASVIEKHFTLDKQMSGPDHKSASLEP 244
QY 257 GELAEILVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVKIPGTILTMDMLTVKV 316
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
245 DEQLQELCTKIRBIESALGDGKQASKSERKNIEIARKSLVAKKKIKKGEIFSEBENLTK- 303
QY 317 GEPKATPPEDIFNL-----VGKKVLVTVEEDTIME 347
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
304 -----RPASGISAMRYDEVILGKASKADYEDDELIHE 334
```

RESULT 14

```
Q8DDZ7 ID Q8DDZ7 PRELIMINARY; PRT; 333 AA.
AC Q8DDZ7;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Sialic acid synthase.
GN Vvi0808.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS016799; AAC09312.1; -.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; NeutB.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; NeutB; 1.
DR PROSITE; PS0844; AFP_LIKE; 1.
KW Complete proteome.
SQ SEQUENCE 333 AA; 36504 MW; 419D1F1A07EBB2A4 CRC64;
```

```
Query Match 27.4%; Score 516.5; DB 16; Length 333;
Best Local Similarity 39.0%; Pred. No. 3.2e-34;
Matches 117; Conservative 53; Mismatches 125; Indels 5; Gaps 3;

QY 20 FIIAETGQNHOGDLDVAKEMIRMAKCGADCAKFKSELEFKNRKALEPPTSKHSW-G 78
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
6 FIIAAGVNHGDIRLAKQLIDAAADAGVDANKFQTKWTELLIVTEDAKMAEVOVENTQRE 65
QY 79 KTYGEHKLHLEFSDHQYRELQRYAEVEGIFFTASGMDMAVEFLHNLVPPFKVSGGDTN 138
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
66 ETQFEMKRLLELSYDDFTLKYCDKGTFTMSTPDEEQSATFDGLQA-VFKIGSGELT 124
QY 139 NPPLYLEKTAAGRPWVSSGMSQMDTMKVQYQIVKPLNPN---FCFLOCTSAYPLOQPDV 195
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
125 NTFPLRHIAFAKPVILSTGMSYLSSEVAHVTLRDAGSLDMLITVLHATTDYPTAPEDV 184
```

```
QY 196 NLRVISEYQKLFDPDIPICYSGHETGIAISVAALGAKVLERHITLDTKWKSGDHSASLE 255
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
185 NLLAMKTEIQAFGIIIVGSDHTLGTETIPVAALGAKVIEKHFTLDTKMTSGPDHKSASLE 244
QY 256 PGELAEILVRSVRLVERALGSPTKQLLPCMACNEKLGKSVVAKVKIPGTILTMDMLTVK 315
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
245 PQELADMTVAIRNIEQALNGWKPVTKEOENRNVKSLVAGKPIWAGSLICADMLEIK 304
```

RESULT 15

```
Q8KNA2 ID Q8KNA2 PRELIMINARY; PRT; 361 AA.
AC Q8KNA2;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Similar to NeutB family.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22053227; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutayavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;
RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa.";
RL J. Bacteriol. 184:3614-3622(2002).
DR EMBL; AF498403; AAM27589.1; -.
DR GO; GO:0016051; P:carbohydrate biosynthesis; IEA.
DR InterPro; IPR006014; Antifreeze dom.
DR InterPro; IPR006190; Antifreeze_like.
DR InterPro; IPR004144; NeutB.
DR Pfam; PF01354; Antifreeze; 1.
DR Pfam; PF03102; NeutB; 1.
DR PROSITE; PS0844; AFP_LIKE; 1.
SQ SEQUENCE 361 AA; 39425 MW; E63D3B9BF5DD76D CRC64;
```

```
Query Match 27.2%; Score 514; DB 2; Length 361;
Best Local Similarity 35.8%; Pred. No. 5.7e-34;
Matches 122; Conservative 55; Mismatches 132; Indels 32; Gaps 5;

QY 20 FIIAETGQNHOGDLDVAKEMIRMAKCGADCAKFKSELEFKNRKALEPPTSKHSWGK 79
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
10 FVIAEAGVNHGDRDLAFKLIDVAAQAGVDVAKFQTFNAKRLASRSAPKANY-QKHETDV 68
QY 80 TYGE--HKRHLRFSDHQYRELQRYAEVEGIFFTASGMDMAVEFLHNLVPPFKVSGGDT 137
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
69 TESQLAMLKLELPKEWHFELQNAHNGIEFISTAFDSDSLAFLAEMQLPFFKVFSGEL 128
QY 138 NNPPYLEKTAAGRPWVSSGMSQMDTMKVQYQIV----- 172
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
129 TNGPLLWAFAKTKGKPLILSTGMATLSEVEGQLAIVAHALSCDNEPKDMDVWRLWNSPV 188
QY 173 -KPLNPNFCFLOCTSAYPLOQPDVNLRVISEYQKLPDIPICYSGHETGIAISVAALG 231
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
189 RMQLQGHVSLHCTSOYPTPPDEVNLLAMDTURSF--GLAVGYSDDHTGELVFIAVARG 246
QY 232 AKVLERHITLDTKWKSGDHSASLEPGELAEILVRSVRLVERALGSPTKQLLPCMACNEKL 291
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
247 ACIIKHFTLDRSMGPDHKSASLEPGELAQWVAQIEMLEVALGSPYKAPQPSWDTRQAA 306
QY 292 GKSVAVKVPIPGTILTMDMLTVKVGEPKAYPPEDIFNLVG 332
Db ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
307 RQCVAAARDIEAGMIITRDDLT-ARSGHGLPPTSLWELVG 346
```

Search completed: September 13, 2004, 14:43:34

Job time : 121 secs